The potential for linked data as a local health intelligence tool for child and maternal health in the UK

Hollie Caitlin Henderson

PhD

University of York
Health Sciences

September 2023

Abstract

Data that are routinely collected when people interact with public services such as health, education, and social care, can be linked together for research. A series of studies are being established across the UK, known as the Born and Bred in (BaBi) Network, which gain consent from pregnant women to use these data about themselves and their child for research. The aim is for these data to inform local decision-making.

In a mapping review, I found limited evidence of linked routine data being used to inform early years decision-making or successful strategies for promoting the use of these data to policymakers. In this thesis, I explore whether linked routine data can be used as a local health intelligence tool for child and maternal health, using the BaBi Network as a case study.

Using linked routine data from the BaBi study in Bradford, I explored whether the available data could be used to address a research question that was prioritised with local stakeholders. I found that key information needed to address this research priority was not available.

Semi structured interviews with local early years decision-makers revealed that although decision-makers perceive value in linked data research, they also have concerns over the quality of routine data and the limitations of the systems that record this information.

My thesis concludes that there is scope to significantly improve routine data sources, which could strengthen their ability to be used to explore local research priorities and inform decision-making. It contributes to a body of evidence demonstrating the challenges of using linked routine data for research and translating this research into policymaking. Future research priorities are to understand if and why key information on local families is missing and to work with local public services and research teams who link data to address these issues.

Table of Contents

List of Figures

List of Tables

List of Boxes

Abbreviations

A&E	Accident and Emergency
ACEs	Adverse Childhood Experiences
ADR	Administrative Data Research
ADRC	Administrative Data Research Centre
ALSPAC	Avon Longitudinal Study of Parents and Children
APEASE	Affordability, Practicability, Effectiveness and Cost-effectiveness, Acceptability, Side-effects/ Safety and Equity
ASQ	Ages and Stages Questionnaire
ASQ-3	Ages and Stages Questionnaires Third Edition
ASQ-SE-2	Ages and Stages Questionnaire: Social and Emotional Second Edition
ASSIA	Applied Social Sciences Index and Abstracts
BaBi	Born and Bred in
BAME	Black and Ethnic Minority
BiB	Born in Bradford
BiB4All	Born in Bradford 4 All
BiBBS	Born in Bradford Better Start
CALLS-Hub	Census and Administrative data LongitudinaL Studies Hub
CDLS	Clinical Data Linkage Service
C-GULL	Children Growing up in Liverpool
CheReL	Centre for Health Record Linkage
CHI	Community Health Index
CHNRI	Child Health and Nutrition Initiative
CI	Confidence Interval
CINHAL	Cumulated Index to Nursing and Allied Health Literature
CLOSER	Cohort and LOngitudinal Studies Enhancement Resources
CMD	Common Mental Disorders
CPP	Child Parent Psychotherapy
CPRD	Clinical Practice Research Database
CRFR	Centre for Research on Families and Relationships
CRIS	Clinical Record Interactive Search
CTV3	Clinical Terms Version 3
DPA	Data Protection Act
EBP	Evidence Based Policymaking
ECHILD	Education and Child Health Insights from Linked Data
e-cohort	Electronic Birth Cohort
eCRUSAD	Early Career Researchers Using Scottish Administrative Data
ED	Emergency Department

eDRIS	eData Research and Innovation Service
ELIXIR	Early Life Cross Linkage in Research
ELP	Early Life and Prevention
EPDS	Edinburgh Postnatal Depression Scale
EPPI-Centre	Evidence for Policy and Practice Information and Co-ordinating Centre
ESRC	Economic Social Research Council
FNP	Family Nurse Partnership
GA	Gestational Age
GDPR	General Data Protection Regulation
GP	General Practitioner
GUS	Growing Up in Scotland
HCS	Hertfordshire Cohort Study
HDR	Health Data Research
HDRC	Health Determinants Research Collaboration
HES	Hospital Episode Statistics
HH	Hollie Henderson
HMIC	Health Management Information Consortium
HMRC	His Majesty's Revenue and Customs
HRA	Health Research Authority
ICB	Integrated Care Board
ICS	Integrated Care System
IMD	Index of Multiple Deprivation
JA	Julie Appleyard
JLA	James Lind Alliance
LHI	Local Health Intelligence (LHI)
LP	Louise Padgett
MatCHNet	Maternity and Child Health Network
MB	Maria Bryant
MCDA	Multi-Criteria Decision Analyses
MCHP	Manitoba Centre for Health Policy
MCS	Millennium Cohort Study
MHRA	Medicines and Healthcare products Regulatory Agency
MIREDA	Mother and Infant Research Electronic Data Analysis
MRC	Medical Research Council
NCCHD	National Community Child Health Database
NCT	National Childbirth Trust
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NICU	Neonatal Intensive Care Unit

NIHR	National Institute for Health and Care Research
NILS	Northern Ireland Longitudinal Study
NNRD	National Neonatal Research Database
NPD	National Pupil database
NTK	Need To Know
OCD	Obsessive-Compulsive-Disorder
OECD	Organisation for Economic Co-operation and Development
ONS	Office for National Statistics
ONS LS	Office for National Statistics Longitudinal Study
OR	Odds Ratio
PARiHS	Promoting Action on Research Implementation in Health Services
PCC	Population, Concept and Context
PhD	Doctor of Philosophy
PICOC	Population, Intervention, Comparison, Outcomes and Context
PMH	Perinatal Mental Health
PPI	Patient and Public Involvement
PSP	Priority Setting Partnerships
PTSD	Post-Traumatic Stress Disorder
R&D	Research and Development
RDS	Research Design Service
REF	Research Excellence Framework
SAIL	Secure Anonymised Information Linkage
SB	Sally Bridges
s.d	Standard Deviation
SES	Socioeconomic Status
SHIP	ScottisH Informatics Programme
SILC	Scottish Informatics Linkage Collaboration
SLS	Scottish Longitudinal Study
SWS	Southampton Women's Survey (SWS)
TFD	Think Family Database
TRE	Trusted Research Environment
UK	United Kingdom (UK)
UKDS	United Kingdom Data Service
UKRI	United Kingdom Research and Innovation
US	United States
VCSE	Voluntary, Community, and Social Enterprise
VICTOR	Making Visible the ImpaCT Of Research
WA	Western Australia
WCHADS	Wirral Child Health and Development Study

WYHCP	West Yorkshire Health and Care Partnership
YHARC	Yorkshire and Humber Applied Research Collaboration

Acknowledgements

First and foremost, I would like to thank Professor Kate Pickett, Sally Bridges and Professor James Wilsdon for your guidance, support, and encouragement throughout my PhD research. I am very grateful to have a supervisory team who have allowed me to grow as an independent researcher yet have always been there to provide advice when needed. Thank you also to my Thesis Advisory Panel member, Professor Maria Bryant, for all your advice and support on my thesis.

I am grateful to the White Rose Consortium for providing me with the opportunity to undertake this PhD and for ensuring I had access to key development opportunities. Without their financial support, this PhD would not have been possible.

A special thank you to all the people who were involved in my prioritisation workshops and public and stakeholder involvement work. Your knowledge and expertise were invaluable to this project. I would like to extend my thanks to the fifteen participants in the qualitative research for giving up your valuable time and sharing your experiences with me. I would also like to thank Born in Bradford for the opportunity to work on the 'Born and Bred in' project and use data from the *'Born in Bradford 4 All'* cohort, which would not be possible without the participants, health professionals, and researchers who make it happen.

I would like to express my gratitude to my colleagues and friends at Born in Bradford and the Public Health and Society Research Group at the University of York, for providing a welcoming and inclusive place to work. To those who have given up their time to discuss or read my work, I am eternally grateful.

Thank you to all my friends, old and new, who have been there for me throughout this PhD journey. A particular thanks to my fellow PhD colleagues Louise Padgett, Kate Mooney, and Jennie Lister for the solidarity you have provided throughout the challenging stages of the PhD. I would also like to thank my oldest friends from school and university, as well as my friends in York, for providing laughter and a much needed distraction from my studies. I am truly grateful to my partner, Robert Malcolm, for your love, kindness, and patience. Thank you for constantly reminding me that I am capable and for never complaining when I wanted to pick your brains.

Last, but not least, I am very grateful to my family. To my mum, for being my biggest cheerleader and always being there for me, and to my sister, for the love and support.

Authors Declaration

I declare that this thesis is a presentation of original work, and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.

Parts of this thesis have been disseminated at three conferences:

- Henderson, H.C., Padgett, L., Bridges, S., Pickett, K.E., and Wilsdon, J., (2023). Society for Social Medicine & Population Health Annual Scientific Meeting. How has linked data research informed early years decision-making? A mapping review. Poster presentation. (Chapter 2)
- Henderson, H.C., Bridges, S., Pickett, K.E., and Wilsdon, J., (2023). ADR UK Conference (2023). What's the big idea? Identifying research priorities using linked data. Oral presentation. (Chapter 4)
- Henderson, H.C., Bridges, S., Pickett, K.E., and Wilsdon, J., (2023). Health Services Research UK Conference. How can we support the use of linked routine data in early years decision-making, Poster presentation. (Chapter 6)

Thesis Outline

This thesis consists of three sections: A) Background (Chapters 1-2), B) Born and Bred in (BaBi) Network Case Study (Chapters 3-6), and C) Discussion (Chapter 7).

In Section A, I provide the background of this research across two chapters. Chapter 1 defines the concepts of interest (linked data, routine data, the early years period) and considers how linked routine data can be used to conduct research around early life health. It also describes the context of health care decision-making and the relevant theories associated with studying the use of research in decision-making. Chapter 2 presents a mapping review of the literature to provide an overall description of how linked routine data have been previously used to inform early years decision-making in the UK. Together, these chapters provide the context and theoretical justification for the research conducted in Section B.

In Section B, I use the BaBi Network as a case study to explore how linked routine data can be used as a local health intelligence tool for child and maternal health. In Chapter 3, I provide a description of the BaBi Network, the research setting and outline the aims and objectives of this thesis. I also present the BaBi Local Health Intelligence model that underpins the research conducted in Chapters 4-6 and justify why this approach is necessary. Chapters 4-6 focus on some of the stages of using linked routine data to inform local early years decision-making: 1) identifying research priorities for linked data research around the theme of early life health; 2) addressing local research priorities using linked data and 3) how local early years decision-makers can be engaged and supported to make use of linked data research.

In Section C, I revisit and discuss the results from all the studies in relation to the overall thesis aims and objectives. I summarise the key findings, the implications of these findings, the strengths and limitations of the case study, and make recommendations for future research, policy, and practice.

This PhD is part of the National Institute for Health and Care Research (NIHR), Yorkshire and Humber Applied Research Collaboration (YHARC). Thus, the methods for this research were chosen to ensure the findings can have a practical application to improving health and care locally.

Section A: Background

This section contains the background for my thesis across two chapters:

- Chapter 1 defines the concepts of interest and the relevant theories associated with the research questions addressed in this thesis. As such, Chapter 1 discusses linked routine data and the potential of using these data to inform decisions around early life health.
- Chapter 2 provides a mapping review of the literature, exploring how linked data research has previously been used to inform early years decision-making in the UK. It aims to identify gaps in the knowledge to inform the research questions addressed in this thesis.

Together, these chapters provide the theoretical framework upon which this research is based.

Chapter 1: Background: defining terms and theories

1.1 Introduction

This section defines key terms and outlines the relevant theories and literature associated with the research questions addressed in this thesis.

Firstly, section 1.2 explains the terms *'routine data'* and *'linked data'* and how these data can be used for health-related research. Section 1.3 defines the early years period and discusses why the early years of life are an important topic for research. It also considers how routine data and linked data can be used to conduct research in this area. Section 1.4 describes aspects of health care decision-making in the United Kingdom (UK) that are relevant to the research conducted in this thesis. Section 1.5 presents a summary of the theories around the use of research by policymakers and the factors that influence this. This chapter concludes with a discussion of how these theories and evidence have been considered in this thesis.

This thesis refers frequently to *'policy'* and *'evidence'*, and it is helpful to clarify the meaning of these terms. Policy can be defined as a plan of action adopted by an individual or organisation such as the government (Collins Dictionary, 2023a). Blum and Pattyn (2022) discuss how policy and evidence can be understood differently depending on the context. In public policy, *'evidence'* is often used to refer to any knowledge or information, although some parts of public policy consider specifically scientific evidence.

1.2 Routine and linked data

Information is collected when people interact with public services in the UK (such as health, education, and social care). This is known as administrative or routine data (ADR UK, 2021). In health care, information on a patient's demographics, lifestyle, clinical diagnoses, engagement with health services, prescriptions, investigations, and medical procedures are predominantly captured in an electronic health record (Goldacre and Morely, 2022). This information is recorded using terms or codes. For example, the National Health Service (NHS) use Read Codes which are a thesaurus of clinical terms (NHS Digital, 2022). The NHS is currently using Clinical Terms Version 3 (CTV3) Read Codes.

Public services often use different databases. This includes health care, where primary and secondary care, midwifery, and health visiting data are stored separately, with lack

of integration between service datasets. Sharing data between public services can allow for a more complete picture of an individual's interactions with these services. This can help to build a picture of the issues faced by communities and families.

The terms '*data linkage', 'record linkage',* or *'linked data'* are used to describe the joining of two or more data sources relating to the same individual, to produce a wealth of information for research purposes (NIHR Applied Research Collaboration South London, 2021; Scottish Government, 2012b). This includes linking routinely collected data across public services to understand how policy initiatives are affecting population health (Tweed *et al.,* 2022).

1.2.1 Benefits of linked routine data for research

The use of existing data is viewed as a cost-effective way of supporting research in public health (Green *et al.,* 2015). There is the potential for linked routine data to be used to:

- Describe a health problem and the extent to which it exists. This could support better identification of need, allowing for early intervention and strategic resource planning.
- Describe and track patterns in health, as well as access to and usage of health services over time.
- Design and/or evaluate an intervention including long-term follow-up of randomised control trials.
- Explore the relationship between health problems and the wider determinants of health, as it can bring together data from across multiple areas that impact health.
- Evaluate the impact of a change to the provision of a service and understand what is effective for different population groups, to better allocate resources (The National Lottery Community Fund, 2022; Ministry of Housing Communities and Local Government, 2021; Franklin *et al.*, 2022; Lewsey *et al.,* 2000; Harron *et al.,* 2014).

The potential benefits of linked data are widely reported in the literature and were noted as early as 1874 by William Farr (Acheson, 1964). Linked datasets can be useful for questions that require large sample sizes, if linkage at the population level can be achieved, and linkage can generate evidence with high external validity (Harron, *et al.,* 2017). Harnessing the power of existing data can also reduce the costs associated with long-term follow-up of participants over time, as well as reducing the burden on participants, as active contribution to the research is not required (Todd *et al.,* 2019).

It is commonly accepted that some of the wider determinants of health lie beyond the health service. This includes a wide variety of factors such as housing quality, level of social isolation, diet quality, level of income, employment status, and access to parks and green spaces (Goldacre and Morely, 2022). Therefore, linking data across health and non-health organisations can facilitate the understanding of upstream influences on health (Sohal, *et al.,* 2022). For example, linked data could be used to investigate if babies who spend time in specialist care units need more support in school. This would involve linking routinely collected health records to education records and the findings could be shared with health and education organisations to inform protective policies (NIHR Applied Research Collaboration South London, 2021).

1.2.2 Challenges associated with linked routine data for research

There are some challenges and drawbacks associated with using linked routine data for research. Routine data are usually collected for administrative purposes, such as delivering health care, and are not intended for research. Hence, data are not always relevant for research and the quality of the data may not be the same standard as research data (Robling *et al.*, 2021). This can result in large investments by researchers into ensuring the quality and usability of these data.

Unpredictable timelines for receiving routine data extracts are also a common challenge faced by researchers applying to use these data. Researchers have described waiting several years for routine data extracts, where they often need to work with each data provider individually (Warren-Gash, 2017). This often leads to projects being abandoned and important research is not conducted (Tweed *et al.,* 2017). A wealth of research explores the challenges of linking data for research and accessing it (Harron *et al.,* 2017; Warren-Gash, 2017; Lugg-Widger *et al.*, 2018; Raftery *et al.,* 2005), however, there has been little focus on the opportunities and challenges of using routine data once they are linked.

Deeny and Steventon, (2015) explain how routine health data can provide only a *'vague shadow'* of a person, meaning that researchers cannot assume that routine data provides an accurate or full picture of a patient's actual experiences. This is because clinical professionals record information relevant to their clinical encounter and mandatory reporting and not necessarily all the information about a patient. Deeny and Steventon (2015) explain that these *'data shadows'* can be interpreted to improve health and care, but that a better understanding of the factors that influence how the data are collected is needed, e.g., understanding why some data are recorded and others are not.

Herzog *et al.* (2007), Harron *et al.,* (2017) and Davis *et al.*, (2016) describe some of the common problems experienced by researchers analysing routinely collected electronic health data. These include:

- Duplication of records, where one individual has multiple records.
- Missing data, which can become a problem if missingness is correlated with health care usage. For example, those who are in poor health are likely to access health care services more regularly, therefore, there are more opportunities to record their information.
- Data can also be missing due to incomplete recording of information by the person inputting the data, or if a person's record has not been able to be linked. For example, a person's education record may be missing due to insufficient identifying information.
- Inaccuracy of data, which can occur when the clinician inputting data fails to find the signs/symptoms of the correct condition or records a diagnosis in the record that differs from their actual diagnosis (diagnostic error). Administrative errors, which occur when turning clinical diagnoses into codes and attaching these codes to the correct record, can also affect the accuracy of the data. This could result in identifying a relationship between two variables that does not exist. These errors can also occur when collecting data as part of other public services.
- Differences in coding practices across data sources, which creates challenges when linking the data together. For example, ethnicity is recorded in primary care and in hospital records, but how it is recorded can differ between the two services.
- Differences in operating systems across data sources.
- Researchers have less information about how these data were linked, since data are often linked by a third party. This can impact the reliability of the resulting dataset.

Furthermore, understanding what the codes and terms used to record the information in each of the datasets mean for research can be a time-consuming process. For example, the code *"pre-diabetes"* could have a wide range of meanings, which depend on the health care setting, and this is important when relying on this for analysis (Goldacre and Morely, 2022). NHS data are recorded at a granular level, where data are collected about diagnostic events, treatment events, symptoms, referrals and more. Hence, to look at the prevalence of a specific disease, researchers may need to access all this detailed data to create a single variable for that disease area.

Using routine data to inform decisions around health can also pose risks to increasing health inequalities. Some groups may be underrepresented in routine data due to having

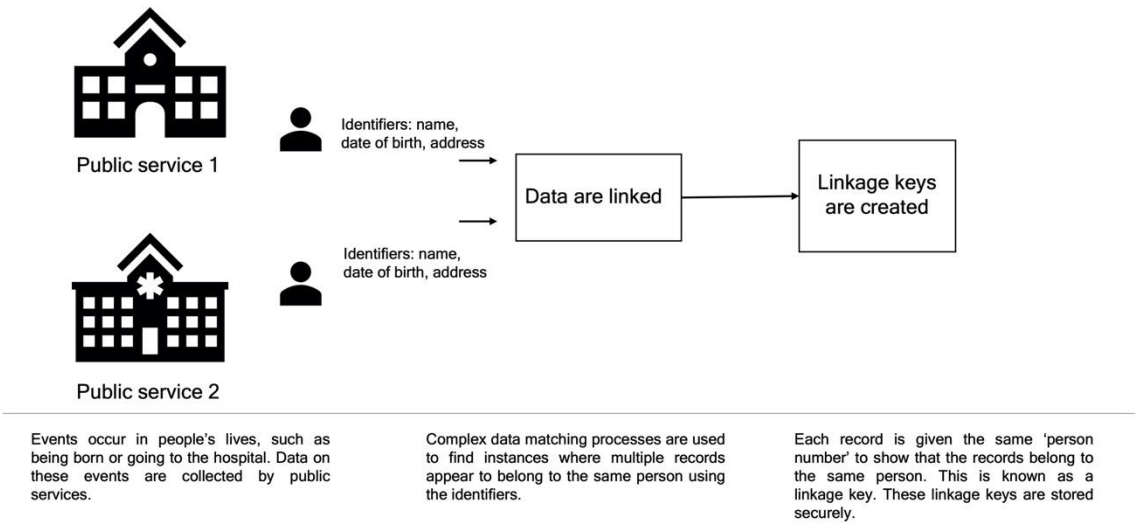
poorer access to health care services relative to their need. Thus, the use of these data may not equitably benefit patients from these groups (Keith *et al.,* 2022).

However, allowing researchers to work with routine data can provide opportunities for researchers to feedback to service providers about the challenges faced. This can contribute to improved clinical reporting in health care (Carson, 2020; ADR UK, 2022a). As health professionals are responsible for generating electronic health care data, they play a key role in ensuring successful use of these data (OECD, 2019). There is also the opportunity for learning to be shared between researchers to improve their understanding of how to use these data for future applications. Although, Warren-Gash (2017) describes lack of engagement between researchers and data providers, which can reduce the opportunities for this feedback loop.

1.2.3 Models of linking routine data

Data can be linked in different ways. In Scotland, every person registered with a general practice is allocated a unique identifying number from a centrally maintained register, the Community Health Index (CHI). This unique patient identifier is also used in hospital based clinical information systems, which can be used to link health data for individuals in Scotland. Similarly, in England and Wales, all persons registered with the NHS, the publicly funded universal health care system, are assigned a unique 10-digit NHS number. This NHS number can be used as a personal identifier across different NHS organisations and used to link data (Carson, *et al.,* 2020). It becomes more challenging to link data across wider public services as there is not a single identifier that is used across the databases. Thus, a combination of identifiers such as name, date of birth, and address must be used to link these data. A set of rules or a criterion can be applied to determine whether the records likely belong to the same individuals (Harron, 2016). In Figure 1, I have summarised how data from two public services, that belong to the same individual, can be linked together (Boyle and Emery, 2017).

Figure 1 A summary of how routine data from two public services can be linked



1.2.3.1 Legal basis and consent for linking routine data

Processing and using personal data for research requires a lawful basis and a number of general legal frameworks exist. Examples include, the General Data Protection Regulation (GDPR), which applies to European Union Countries and the UK, and the UK Data Protection Act (DPA) 2018 (Regulation (EU) 2016/679, 2016).

In most cases, the lawful basis for research-related data processing is either:

- Public task, where the processing of data is necessary for performing a task in the public interest (Article 6(1)(e) in the GDPR); or
- Legitimate interests, where the processing of data is necessary for legitimate interests (ICO, 2023).

Where personal information is sensitive, the legal basis can be (Article 9(2)(j)):

- Archiving purposes in the public interest,
- Scientific or historical research purposes; or
- Statistical purposes.

The GDPR defines *'sensitive personal information'* as information that can reveal a person's racial or ethnic origin, political opinions, philosophical beliefs, trade union membership, as well as genetic data or biometric data, and data concerning health sex life or sexual orientation. Hence, routine data can be considered sensitive personal data.

Schedule 1 paragraph 4 of the DPA 2018 sets out additional requirements for sensitive data including that the processing of data is necessary for that purpose; it is subject to

appropriate safeguards; is not likely to cause someone substantial damage or distress; is in the public interest, and it is not used for decisions about particular people, with exception of approved medical research (ICO, 2023).

Thus, the legal basis for processing and linking routinely collected data is complex and can vary depending on the context. Studies that link routine data for research often set out the legal basis for processing these data in a privacy notice on their website.

Consent is often sought from participants to take part in research studies, including those linking routine data. It is an important ethical standard that protects the autonomy and privacy of participants in research studies. Consent to take part in research is separate from the lawful basis to process personal data, however, it is often collected to help ensure research is ethical and complies with other applicable laws.

The two models of consent that are often referred to by data linkage studies are: the opt-in model and the opt-out model of consent. An opt-in model of consent typically requires a person to actively sign up for their data to be collected and used. An opt-out consent model means that a person's data will be collected and used automatically, unless the person actively withdraws their consent (Understanding Patient Data, 2018).

Under opt-out consent models, relatively fewer people actively opt-out of their data being used for research, which can lead to a higher coverage of the population in a dataset. This is important for a diverse population (Understanding Patient Data, 2018). Under opt-in consent models, data are more likely to reflect a select portion of the population who actively agree to share their data, which can result in selection bias (De Man *et al.,* 2023).

1.2.4 Concerns related to data privacy

Data privacy is important when using routinely collected data for research. Concerns related to the privacy of individuals have been the downfall of previous NHS data sharing projects. For example, the care.data programme was a national data sharing initiative that aimed to develop a central database linking NHS hospital and general practice data for more efficient health care and research (Vezyridis, 2017). However, it failed to win the public's trust and lost the support of doctors because of frequent data breaches and because the benefits of data sharing were not adequately promoted (Godlee, 2016). Thus, large numbers of patients opted out of having their data shared, which compromised the usefulness of these data. This was despite having a robust legal basis in the Health and Social Care Act 2012 and compliance with the current data protection law. This demonstrates the importance of public trust and having a system that takes credible steps to protect privacy (Goldacre and Morely, 2022). Since then, research has shown that members of the public in the UK are broadly supportive of research utilising

linked routine data, providing there are appropriate safeguards in place (Waind, 2020; Goldacre and Morely, 2022; Kalkman *et al.,* 2022). In addition, legislation such as the Digital Economy Act 2017 provides a legal gateway for non-health data held by public authorities to be made available for research.

1.2.5 Accessing and using linked routine data

Researchers can apply to access routinely collected health datasets in the UK for research and innovation, although there are strict controls in place to monitor the use of these data. For example, researchers are only given access to the minimum amount of data necessary to conduct the research (Health Data Research, 2023).

The term data custodian is used to describe an individual or organisation who controls how data are accessed and used for research. The data custodian is responsible for ensuring that any processing of personally identifiable data is safe and lawful (Health Data Research, 2023). Health Data Research UK (HDR, UK) designed the *'Five Safes Framework'* to help data custodians make decisions that will enable effective use of confidential and sensitive data. The five dimensions are summarised here (Health Data Research, 2023):

- **Safe People:** Can the data user be trusted to use the data appropriately?
- **Safe Projects:** Is the use of these data appropriate, lawful, and ethical? Will the project deliver public benefit?
- **Safe Settings:** Are there appropriate controls in place for accessing the data?
- **Safe Data:** Has the data been treated appropriately to minimise the risk of identification of individuals or organisations?
- **Safe Outputs:** Do the results of the research that uses these data prevent identification of individuals from the data?

Once a researcher's application to use these data has been approved, a legal contract must be signed, known as a *'data sharing contract'*. This contract sets out clear rules such as how the data will be provided; the purpose for which the data can be used and any restrictions; how and when data must be destroyed after use; and data security requirements (Health Data Research, 2023). It is also important that data are stored securely and there are robust IT systems in place to support this.

Before data are shared and used for analysis, data are pseudonymised. This means all information that could directly identify an individual is removed and replaced with an unrelated unique identifier (Goldacre and Morely, 2022). Examples of directly identifiable information include name, date of birth, and address of the patient or person. Even with

pseudonymisation, there are still risks that an individual could be identified, especially when datasets cover a large proportion of the total population. Hence a system needs to be put in place to protect these data.

Currently, many data projects exist, where the same patient records are being distributed to each of the project teams. This is inefficient as it duplicates implementation and data preparation costs, whilst also obstructing the sharing of ideas between settings (Goldacre and Morely, 2022). Once researchers have used these datasets, they are then securely destroyed. In April 2022, an independent review by Ben Goldacre outlined how to achieve efficient and safe use of health data for research and analysis (Goldacre and Morely, 2022). Box 1 discusses the key points from this review that are relevant to this thesis.

Box 1 Key points from the Goldacre Review

The Goldacre review was commissioned by the Secretary of State for Health and Social Care in the UK to inform and accompany the NHS Data Strategy. The focus was on facilitating access to NHS data by researchers, commissioners, and innovators, where patients can opt-out of having their data shared. The review recommends investment in a coherent approach to data curation and the development of a number of secure platforms to unlock the potential of NHS data and maintain patient privacy. These are known as Trusted Research Environments (TREs).

A TRE is a secure platform, where data flows from multiple sources and researchers can come to work on these data remotely. Researchers only export the answers to their analytical questions and data stays in a secure environment. This centralised model presents an opportunity for data owners to make extracts of data available to multiple researchers, securely, and allows researchers to contribute to the development of these data as an ongoing resource for research (Mc Grath-Lone *et al.,* 2022). It prevents data being transferred to individual researchers who are then responsible for storing, analysing, and destroying the data. This allows for greater transparency as usage of these data can be monitored (Goldacre and Morely, 2022). The government have acknowledged this recommendation by investing £200 million in the development of TRE's for improved data access for research (Department for Business Energy & Industrial Strategy & Department of Health and Social Care, 2022).

1.3 Importance of early years of life

In the context of this thesis, the *'early years of life'* are defined as the period from conception to age five years. The early years of life are a critical period for child health and development (Early Intervention Foundation, 2021; House of Commons and Health and Social Care Committee, 2019). What happens during this time can have lifelong consequences on health outcomes including obesity, heart disease, and mental health, as well as effects on other outcomes such as educational attainment and economic status (Marmot *et al.,* 2010). The UK government spends billions of pounds on addressing these challenges, some of which could be prevented by focusing on addressing the issues that occur in the early years period (Department of Health and Social Care, 2021b).

Specifically, leading child health experts have agreed that care given in the first 1,001 days of life has more influence on a child's future than care provided at any other time in their life (Department of Health and Social Care,2021b). During these early years, infants are susceptible to their environment and are completely reliant on their caregivers. Therefore, it is important that parents receive the right support during this time. If developmental delays are not identified and addressed early on, this could cause problems later in life (Department of Health and Social Care, 2021b). Hence, the World Health Organisation describes investment in the early years of life as *"one of the greatest potentials to reduce health inequities within a generation"* (World Health Organisation, 2008 p.4).

In the UK, early years' service providers such as midwives, health visitors, primary care practitioners, dentists, social workers, and committed volunteers play a critical role in child and maternal health. They offer a broad range of services to families, as well as more targeted services in response to need (Department of Health and Social Care, 2021b; Bryar *et al.,* 2013). These services can include breastfeeding support, mental health support, smoking cessation, and parenting support. Box 2 provides more detail on the role of UK midwifery and health visiting services in supporting children and families (Department of Health and Social Care, 2021b).

Box 2 Early Years Services

Midwifery

Midwives are registered health professionals who provide support to mothers throughout pregnancy including during labour, the birth of the baby, and up to ten days after the baby is born. Midwives ensure that parents and carers feel prepared for the baby's birth. Midwife appointments are a universal service provided by the NHS.

Health visiting

Health visitors provide support after the baby is born up until the child starts school. They are registered nurses who visit every family with a new baby and focus on prevention and promotion of health. Health visitors play a vital role in reducing long-term health inequalities and prevention of long-term conditions as well as providing support with immediate parenting concerns. They offer a minimum of five health and development reviews to parents. Parents first meet their health visitor for an antenatal review when they are 28 weeks pregnant to discuss their health and the transition into parenthood. The health visitor then offers a review usually at ten to fourteen days after the baby is born, at six to eight weeks, nine to twelve months, and when the child is aged 24 and 30 months. Local authorities commission health visiting services as part of the Healthy Child Programme.

The importance of investing in the early years, as a means of improving health across the life course, is being increasingly recognised by policymakers. The use of research evidence to underpin decision-making regarding the early years is strongly promoted (Mikkelsen, et al., 2019; Smith, 2013; Jacob *et al.,* 2019; Orton, *et al.,* 2011; House of Commons and Health and Social Care Committee, 2019). In the UK, the Department for Health and Social Care published '*The Best Start in Life: a vision for the 1,001 critical days'*, in March 2021 (Department of Health and Social Care, 2021b). It was developed as part of the early years' healthy development review and outlines six areas for action to improve health outcomes of all babies in England. In addition, the NHS Long-term Plan emphasises the importance of giving every child the best start in life (NHS, 2019).

1.3.1 Using linked data for research into the early years of life

Much of the evidence underpinning the idea that the early years are important, stems from large birth cohort studies. Birth cohort studies follow babies, and usually mothers, over time to see how a combination of factors including lifestyle, genetics, and the environment might impact on long-term health (Wright *et al.*, 2013; Boyd *et al.*, 2013;

Syddall, et al., 2019). They collect sociodemographic information on parents and children, as well as data on key outcomes and exposures to allow life course research to be conducted. Therefore, key research hypotheses need to be considered at the cohort's inception, as it is crucial that the right data items are captured (Lawlor *et al.,* 2009). However, collecting bespoke data on thousands of participants over time can be expensive and burdensome to participants. These studies are exposed to risks such as participants lost to follow-up and selection bias (Carson, *et al.,* 2020). Their responsiveness to emerging threats such as climate change and Covid-19, can be limited, as variables required to explore these may not have been included at the offset.

There is the opportunity for linked routine data, that's collected by public services over the course of people's lives, to be used to explore meaningful relationships between the early years of life and health outcomes later in life. The use of linked data for early years research can be demonstrated in countries with more established data linkage systems than the UK. Hence, valuable lessons can be drawn about how linked routine data can be used to support research into the early years of life.

Linking birth registers with administrative health records for mothers and babies was first achieved in Nordic countries, where they have established a nationwide infrastructure for collecting and linking data for all citizens (Olsen *et al.,* 2001). Citizens of these countries carry a unique personal identity number that allow registers using personal identity numbers to be linked together. The Nordic countries have also kept medical birth registers, containing basic information on the mother, father, and new-born child, which can then be linked to other national databases (Langhoff-Roos, 2014). Women who gave birth at the time when the registers were established have now reached 60-80 years old, meaning relationships between pregnancy and subsequent health outcomes can be examined. Thus, routine data have been used to increase knowledge of certain risk factors in pregnancy and have resulted in changes to national guidelines (Langhoff-Roos, 2014). The success of these data linkage systems is underpinned by wide public acceptance that individual information is important for research. The public also assume that personal integrity is protected and appropriate safeguards, such as anonymity, are in place (Langhoff-Roos, 2014).

Cross-sectoral data linkage in Australia is also well established. The Western Australia Developmental Pathways project, founded in 2005, links a range of routinely collected data to investigate factors influencing differences in developmental outcomes for children (Telethon Kids Institute, 2021).

The Centre for Health Record Linkage (CHeReL) is a data linkage facility connecting data from New South Wales and Australian Capital Territory, who each have their own governments and Health and Human Service systems (Irvine, *et al.,* 2019). These linked data have been used to validate the accuracy of screening tests including antenatal serum screening and how this links to pregnancy outcomes (CHeReL, 2021).

A final example is the Early Childhood Health and Development Project, established in 2009. The project aimed to link routine data for a cohort of children born between 1999 and 2011 in South Australia. It involves linking 12 government administrative datasets, including health and education data. Research using these data has been instrumental in building cross-government buy-in that child protection is everybody's responsibility (Pilkington *et al.,* 2019). The project's success is attributed to the positive relationships that have developed between researchers and the individual data custodians (Better Start Child Health and Development Research Group, 2014).

In the UK, data on children's health is held across a number of systems. For example, Child Health Information Services keep records of the child's immunisations, the Maternity Services Dataset holds information about the birth and primary and secondary care captures information on any treatments they may have received. In recent years, there has been significant investment by policymakers into developing systems that facilitate data linkage to support children and families. Box 3 provides examples of these large investments.

Box 3 UK investments in data linkage

- In 2012, the British Health Foundation joined a consortium of UK government and charity funders, led by the Medical Research Council (MRC) to invest £19 million in four electronic health research Centres of Excellence (e-health centres) in London, Manchester, Dundee, and Swansea. A further £20 million investment established the Farr Institute of Health Informatics Research. This investment aimed to support safe use of patient data for medical research, linking information in NHS health records with other forms of research and routinely collected data, for medical research across all diseases (British Health Foundation, 2012).
- In March 2021, The Ministry for Housing, Communities and Local Government, made £7.9m of funding available, known as the *'Local data accelerator fund for children and families',* for innovative local data sharing projects to support children and families. The funding was awarded to projects that use data from different agencies to improve services and support a partnership between the local authority and local services. The funding stream also encouraged areas with current linkage systems to support other local areas to improve their data use, allowing for wider analysis (Ministry for Housing, Communities and Local Government, 2021).
- Administrative Data Research (ADR) UK began as a pilot programme in 2018 to explore different ways of working with data owners and researchers. Following on from the success of this pilot, it has been awarded £105 million of long-term funding up to 2026 (ADR UK, 2022b).
- The National Lottery Fund invested £215 million to improve the life chances of babies in five local areas in England. They are leveraging routine data to collate information on key outcomes to evaluate the impact of their investment (The National Lottery Community Fund, 2022).

A number of data linkage studies, which include data on the early years of life, exist in the UK at the local, regional, and national level. There are studies established with the intention of using and linking predominantly routinely collected data, such as Born in Bradford 4 All (BiB4All), eLIXIR, and Connected Yorkshire (NIHR Guy's and St Thomas' BRC, 2023; Born in Bradford, 2023, Bradford Institute for Health Research, 2023). Some studies link data across public services, whilst others link across health datasets.

Numerous established UK research cohort studies have now sought consent to link routinely collected data to bespoke data collected by the longitudinal study, including Avon Longitudinal Study of Parents And Children (ALSPAC), Born in Bradford (BiB),

Growing Up in Scotland (GUS,) and the Millennium Cohort Study (MCS) (O'Neill, *et al.*, 2019; Raynor, *et al.,* 2008).

There are also *ad hoc* studies which have linked two or more datasets with the intention of answering a specific research question. An example of this is a study by Robling *et al.,* (2021) which uses routine data linkage to follow up on an earlier randomised control trial examining the effectiveness of an intensive home-visitation programme for vulnerable teenage mothers in the UK.

There has also been significant investment into services that facilitate secure and ethical access to linked data for research, which are often referred to as *'safe havens'* or *'TREs'.* An example of a safe haven or TRE is the Secure Anonymised Information Linkage (SAIL) Databank (NIHR Applied Research Collaboration South London, 2021). SAIL provides an infrastructure for secure management and processing of sensitive and confidential data. SAIL also meets the Five Safes, set out in section 1.2.5.

Moreover, the Mother & Infant Research Electronic Data Analysis (MIREDA) partnership, launched in April 2023, brings together leading researchers from Swansea University, the University of Edinburgh, Kings College London, the University of Nottingham, the University of Birmingham, and the Bradford Institute for Health Research, to accelerate capabilities and research on maternal and infant health. It aims to develop new resources and tools for research that use routinely collected data (University of Birmingham, 2023).

A summary of UK data linkage studies, relevant to early life health, is presented in Table 1. This list aims to illustrate the vast array of data linkage studies in the UK and includes the studies that I was aware of at the time of completing this PhD. Table 2 lists some of the services in the UK that link data for research.

Table 1 A summary of UK studies linking routinely collected data

Study	Local or National	Linked /aiming to link	Summary	References
Born in Bradford (BiB); Born in Bradford Better Start (BiBBS); Born in Bradford 4 All (BiB4All); and Born and Bred in (BaBi) Cohorts	Local	GP, Dentists, Midwifery Health Visiting, Department of Work and Pensions, voluntary organisations, local authority, social care, education, children's centres, NHS digital, other research studies, and national child measurement programme.	The BiB longitudinal birth cohort study, launched in 2007, has recruited over 13,500 children, and their parents, who were born at Bradford Royal Infirmary between March 2007 and December 2010. Detailed information on participants was obtained in the form of a questionnaire and permission to link data that has been routinely collected about them. The BiBBS cohort is an experimental birth cohort study evaluating the impact of multiple early life interventions, which links bespoke research data to routine health and local authority data. The BiB4All and BaBi cohorts build on the success of BiB. They gain consent from pregnant women to access and use the data that is routinely collected about themselves and their child for research purposes.	(Born in Bradford, 2019; Born In Bradford, 2021; Born in Bradford, 2023 Dickerson *et al.,* 2016 Wright et al., 2013; Wright *et al*. 2021)
Children Growing up in Liverpool (C-GULL)			C-Gull is a prospective cohort study of 10,000 first-born infants and their parents, that is nested within a population-wide, civic data linkage platform. The study will collect biological, biometric, sociodemographic, and psychosocial information antenatally, at birth and when the child is 3, 12, and 24 months. This will be linked to routine health and education data. Pregnant people, in their first pregnancy, who are over 16 years old and are booked for their maternity care to be provided by Liverpool Women's NHS Foundation Trust are eligible to take part.	(University of Liverpool, 2023)
Connected Yorkshire	Local	Primary care, secondary care, community care and social care	Connected Yorkshire links routine data for 700,000 individuals at Bradford Teaching Hospitals NHS Foundation Trust.	(Bradford Institute for Health Research, 2023)
The Hertfordshire Cohort Study (HCS)	Local	Historical ledgers by midwives and health visitors entered into the NHS central register at Southport linked to	HCS is a cohort of approximately 3000 men and women born in the English County of Hertfordshire during the period 1931-1939 and were still residents in the county in the 1990s. The cohort was introduced and is maintained by the MRC Lifecourse Epidemiology Unit, University of Southampton. The aim of the study is to evaluate the relationship between early growth, in the pre and postnatal stages, genetic factors, adult lifestyle and risk of age-related disorders. The study used	(CLOSER, 2018b)

Study	Local or National	Linked /aiming to link	Summary	References
		hospital admissions and mortality data.	historical ledgers by midwives and health visitors, stored in the NHS central register at Southport (now NHS Digital), to trace those whose birth and infancy data were recorded between 1931 and 1939. This information included birth weight, weight at one year, the method of infant feeding and details of childhood illnesses up the age of five years. Between 1998 and 2004, those who were still alive in Hertfordshire were contacted, with permission for their GP, to be followed-up. This resulted in 2997 men and women attending a clinic for detailed physiological investigations. Since 2004, cohort members have consented to take part in various other follow-up studies, including permission to link their information to hospital admissions and mortality data.	
Growing Up in Scotland (GUS)	National	Health, study data, education.	GUS launched in 2005 and has collected information on three nationally representative cohorts. Birth Cohort one recruited 5,217 children born in 2004/5, Child Cohort two has 2,858 children that were born in 2002/3 and Birth Cohort two recruited 6,127 children, born in 2010/11. In 2017, they invited a further 1,500 children born between June 2004 and May 2005 to be interviewed alongside birth cohort one. Overall, data on around 14,000 children has been collected as part of GUS and has been able to provide vital evidence for the Scottish government, informing policies for children and families. GUS also intended to be a broader resource for academics, voluntary sector organisations, and other interested parties. The study is funded by the Scottish government and is carried out by ScotCen Social Research in collaboration with the Centre for Research on Families and Relationships (CRFR) at the University of Edinburgh, and the MRC Social and Public Health Sciences Unit in Glasgow. GUS has gained consent from parents to link data collected about themselves and their child to routine data by health and education authorities. As part of GUS, families have been regularly visited by trained interviewers to capture information on a range of topics, thus, linking data enhances data they are able to collect.	(Growing up in Scotland, 2021; Growing up in Scotland, 2012)
Office for National Statistics Longitudinal Study (ONS LS)	National	Links census data, life events data (births, deaths, marriages), cancer registrations	ONS LS links data for a 1% sample of the population of England and Wales allowing records of over 500,000 people to be assessed at each point in time. These data have been useful for examining the environmental effects on health and inequalities in health. Initially, data were extracted from the 1971 census and linked to administrative, vital events and health datasets. The dataset is supported by CeISIUS at UCL, see Table 2. CALLS-HUB provides a list of policy briefs	(Office for National Statistics, 2021)

Study	Local or National	Linked /aiming to link	Summary	References
			demonstrating the impact of the ONS LS research, however, there was little evidence of impact in early years policy.	
Scottish Longitudinal Study (SLS)	National	Census data from 1991 onwards; vital events data (births, deaths, marriages); NHS Central Register data (Information services division Scotland); and education data	SLS a replica of the ONS LS and comprises of 274,000 individuals, selected based on 20 random birth dates. Data have subsequently been linked to vital events, school data, weather data, pollution data, and census data from 1991, 2001 and 2011. SLS has undertaken multiple projects relating to child and early life health ranging from fertility to chronic childhood illness and education. The linked data are maintained as a single databank which is accessible on a project-by-project basis	(Scottish Longitudinal Study Development & Support Unit, 2021; Wellcome trust, 2015)
Northern Ireland Longitudinal Study (NILS)	National	Census data, Land and property services data, GP prescribing, vital events, education	Its sample of 500,000 represents approximately 28% of the population of Northern Ireland. The focus of this study is on health-related research including research on migration, fertility, and inequalities.	(Northern Ireland Longitudinal Study Research Support Unit, 2021)
Millennium Cohort study (MCS)	National	Cohort study data, NHS birth and maternity records, emergency department datasets, hospital inpatient datasets, GP data, cancer register and death register	MCS, also known as *'Child of the New Century'* to the cohort members, follows the lives of around 19,000 young people who were born between 2000 and 2002 across the UK. The Centre for Longitudinal Studies at University College London conducts the study. MCS sought written parental consent to link routine data, including health and education, to survey responses for cohort participants when they were aged seven up to their 14th birthday, in which 90.7% of parents consented. At the age of 17, cohort members were asked for their own consent. Initially, when cohort members were approximately nine months old, parental consent was gained to link MCS data to NHS birth records and maternity episode hospital records. Linkage applications were approved for Wales and Scotland but declined for England after two years due to concerns over the wording of consent forms. The resulting linked datasets have been used to examine issues such as childhood obesity, asthma, and patterns in health service utilisation.	(Centre For Longitudinal Studies, 2020; Tingay, 2019)
Generation Scotland	National	GP records, questionnaire data, genetic samples.	Generation Scotland is a research study that explores the health and wellbeing of volunteers and their families by inviting participants to complete questions and provide genetic samples, where these are then combined with NHS health records to provide a rich evidence base for understanding health in Scotland. One of the datasets Generation Scotland has available is the Scottish Family Health Study which is a family-based cohort investigating the genetics behind common complex	(The University of Edinburgh, 2021)

Study	Local or National	Linked /aiming to link	Summary	References
			diseases and responses to treatments. The study recruited between 2006 and 2011 and has over 24,000 participants who have consented to record linkage and recontact.	
Maternity and Child Health Network (MatCHNet)	National	Linked maternal and child data for England, Northern Ireland, Scotland, and Wales. The potential for cross-sectoral linkage to prescriptions, primary care, census, hospitalisations, benefits, education, and housing.	MatCHNet is developing a multidisciplinary community of public health researchers, methodologists, policymakers, and service providers to prioritise upstream policy interventions that can be evaluated using linked administrative data. The network identifies three key life periods where policy intervention can impact maternal and child health, resulting in improvements to noncommunicable diseases: pregnancy, infancy (birth to one year), and early childhood (one to six years). MatCHNet identify the challenges in identifying policy priorities that best address the problems in all four nations and aim to ascertain what administrative longitudinal data can be linked across the four UK countries.	(MatCHNet, 2021)
Avon Longitudinal Study of Parents And Children ALSPAC	Local	Health records, Education records, economics, employment, and social support (benefits) records, criminal convictions and cautions, neighbourhood data, vital life events data, ONS mortality data, and cancer registrations.	ALSPAC, also referred to as *'children of the nineties'*, launched in 1991. Pregnant women who resided in the City of Bristol in the Southwest of England and were expected to deliver their baby between 1st April 1991 to 31st December 1992, were recruited into the study. Subsequently, additional recruitment took place when the children reached seven years old, to capture those that had failed to join the study originally. This resulted in a sample size of around 15,000. The Project to Enhance ALSPAC through Record Linkage was established to enrich the ALSPAC cohort data through linking routine sources of health and social data. The ALSPAC cohort has contributed to the debate on the consumption of fish during pregnancy and has influenced both UK and United States guidance by showing that the benefits of fish consumption outweigh the potential harm. ALSPAC also negotiate access to bespoke linkages. ALSPAC have produced over 2000 publications, many of which have exploited data linkage.	(Audrey *et al.*, 2016)
Early Life Cross Linkage in Research (eLIXIR)	Local	Incorporates clinical data from maternity, neonatal, mental health records, GP, and biological samples. The ambition is to supplement this with	The eLIXIR partnership, developed in 2018, created a repository of real-time, pseudonymised data from electronic health records of two acute and one Mental Health Care NHS Provider in South London. The partnership is funded by the MRC and is a multidisciplinary academic collaboration which aims to provide information on a large number of mothers and babies in a single source, which is naturally accumulating. Data from maternity, neonatal and maternal and child health mental health records are linked together to support research into the early	(Carson, *et al.,* 2020)

Study	Local or National	Linked /aiming to link	Summary	References
		other local and national sources such as NPD and pollution data.	life origins of physical and mental health. Samples collected from women attending antenatal care will provide a biobank, which can provide information on both common and rare complications during pregnancy, neonatal life, and the longer-term health consequences for both the mother and the baby. This partnership has been facilitated by the King's Health Partners. eLIXIR has plans to expand their data linkage model to other geographical settings in the UK with the intention of developing a national data network.	
Think Family Database (TFD) by Bristol insights group			TFD, set up by Bristol city council, links data from 30 different public sector organisations to create a rich dataset covering 54,000 families across the city of Bristol. It was created in response to the national Troubled Families programme, aiming to improve outcomes for families by better targeting services. It enables the identification of vulnerable children and families so that resources can be allocated more effectively.	(Ministry of Housing, Communities and Local Government, 2021)
Understanding Society, The UK household longitudinal study	National	Education, health data, economic data (Department for Work and Pensions, HMRC), geographical data other government departments, and regulatory bodies.	The study started in 2009, building on the British Household Panel Survey. It covers all regions and nations of the UK and collects information annually through a representative sample of 40,000 households of the UK population. Information is collected directly from everyone over the age of 10 years. As part of the survey, parents and carers answer questions about the young children in their care, which cover a range of parenting and child development areas, including pregnancy. By 2020, 27 years of data were available for a significant subgroup part of the British Household Panel Survey. The data have been linked, with consent, to geographical and administrative data, to build up a richer picture of households. Information from the study has been used by researchers to investigate how changes in economic, social health events effect individuals and communities and evidence has been used extensively by government departments, devolved administrations, agencies, charities, and think tanks.	(CLOSER, 2018d; ESRC,2018).
Next steps, previously known as the Longitudinal Study of Young People in England	National	Cohort data, NPD, education, economic data, criminal databases, health data.	The study began in 2004 and follows the lives of 16,000 people born between 1 September 1989 and 31 August 1990. The study has collected a range of information on education, employment, emotional and physical health, and wellbeing. They have gained consent, from cohort members over 25 years, to link data to a wide range of sources.	(ESRC,2018).

Study	Local or National	Linked /aiming to link	Summary	References
National Neonatal Research Database (NNRD)	National		The NNRD collects population data for England for infants born alive between 25 and 31 weeks of gestation as well as 70% of those born at 23 weeks and 90% at 24 weeks. It supports new-born health services. The Neonatal Data Analysis Unit was established in 2007 at Chelsea and Westminster Hospital campus of Imperial College London with the aim of improving the quality of electronic clinical data and promote the use of this data to support neonatal services. All Neonatal Units in England, Wales and Scotland currently contribute data and form the UK Neonatal Collaborative. NNRD data originate from information entered by clinicians (usually trainees) and nursing staff onto the Bager.net platform at the point of care. The database offers the opportunity for collaboration and has been linked with HES to create birth cohorts.	(Battersby, 2018)
ADR Birth Cohort Data Linkage Study	National	93% of birth registration and notification data from 2005 to 2014 with HES, delivery records.	This study is led by Alison Macfarlane at City, University of London. The findings from this study have suggested that births in England and Wales follow a pattern, where the number of births is lower at weekends and on public holidays. These findings have been used to help plan staffing for maternity services to ensure a safe 24hr NHS service. This linked birth cohort is being used for other research such as The Tracking the Impact of Gestational Age on Health, Educational and Economics outcomes: A Longitudinal Record Linkage Study (TIGAR).	(ADR UK, 2021a)

Study	Local or National	Linked /aiming to link	Summary	References
Liverpool Families Programme	Local	Local authority, the Police National Computer and Prisons Database, NPD and Individualised Learner Record, The Work and Pensions Longitudinal Study, HES, Mental Health Minimum Dataset, Improving Access to Psychological Therapies and Maternity and Children's Data Set, National Drug Treatment Monitoring System, Crime Mapping Database.	Integrates 35 streams of data from children's social services, schools, the criminal justice system, health, and benefits data to identify those who are most vulnerable and could benefit from early intervention. The process of data linkage involved local authorities signing up to share personal information about families that were eligible for the Liverpool Families Programme, this data was then sent to the ONS, an appointed third-party processor, who matched the data against national administrative datasets held by government departments and bodies.	(Liverpool City Council, 2021a; Liverpool City Council, 2021b; Ministry of Housing, Communities and Local Government, 2021)
Wirral Child Health and Development Study (WCHADS)	Local	Questionnaires, observational measures, health visitor routine data and teacher reports.	Funded by the MRC, WCHADS was established in 2007 with the aim of identifying why some children develop behavioural problems from an early age. The data collected can be used to examine the earliest origins of child and maternal outcomes. First time pregnant women, aged 18 years and above, were recruited at their 20-week scan appointment by the sole provider of antenatal care in Wirral, Merseyside, between March 2007 and December 2008. WCHADS links their longitudinal data with health visitor routine data and teacher reports.	(UKRI, 2021; CLOSER, 2018c; University of Liverpool, 2021)
Education and Child Health Insights from Linked Data (ECHILD) database	National	NPD and HES	The ECHILD Database will bring together information about health, education, and social care for all children in England to better understand how education affects children's health and the reverse relationship, to provide support to children where needed.	(ADR UK, 2021b)
Abbreviations: Hospital Episode Statistics, HES; General Practitioner, GP; Office for National Statistics, ONS; National Health Service, NHS; Medical Research Council, MRC; National Pupil Database, NPD; His Majesty's Revenue and Customs, HMRC; United Kingdom Research and Innovation, UKRI				

Table 2 A summary of UK data linkage services

Service	Linked data	Summary	References
Clinical Practice Research Datalink (CPRD) - National	HES admitted patient care, HES outpatient, HES accident and emergency, HES diagnostic imaging dataset, ONS death registration, National Cancer Registration, and Analysis Service data from Public Health England including: Cancer registration data, cancer patient experience survey data, systemic anti-cancer treatment data, national radiotherapy dataset, mental health dataset and measures of relative deprivation.	CPRD is a UK government research service that provides anonymised linked national electronic health records to support large scale public health research, whilst protecting patient confidentiality. It encompasses 60 million patient records. It is supported by the MHRA and the NIHR. CPRD operates an opt-out system where GPs' can choose whether to contribute de-identified patient data to CPRD for patients who have consented to their patient record being shared with CPRD or NHS Digital.	(CPRD, 2021; Harron, *et al.,* 2012; Padmanabhan, 2019)
Secure Anonymised Information Linkage (SAIL) databank	Routinely collected national health data and social care datasets including hospital, clinics, and GPs in Wales. Data sources also include non-health data such as social services, housing, transport, and education.	SAIL Databank, established in 2006, is funded by the Welsh Government's National Institute of Social Care and Health Research. SAIL has been able to successfully link over two billion records from various health and social care service providers.	(Lyons *et al.*, 2009)
Clinical Data Linkage Service (CDLS)	HES, neonatal data, national cancer registry, Lambeth DataNet, NPD, ONS Mortality, ZTAS Clozapine Monitoring Service, ONS Census 2011, The eLIXIR Project.	The CDLS enables information from the Clinical Record Interactive Search (CRIS) to be linked with other clinical sources within a secure *'safe haven'*.	(NIHR Applied Research Collaboration South London, 2021)

Service	Linked data	Summary	References
Qresearch	Linked HES and National Death Register data supplied by NHS Digital; Linked Cancer Registry data and Public Health England Second Generation Surveillance System Covid-19 data; Linked Intensive Care National Audit & Research Centre Case Mix Programme data; Cancer Registry Data	Qresearch is a large, consolidated database made up of anonymised health records for over 35 million patients. Linked data are only available on the servers at the University of Oxford.	(Qresearch, 2021)
The Cohort and Longitudinal Studies Enhancement Resources (CLOSER) consortium.		Established in 2012, the CLOSER consortium brings together world-leading longitudinal studies to maximise their use and value to population research. They facilitate linkage of government data with survey data collected by longitudinal studies and have developed a series of work packages to promote good data linkage practices and improved access to linked data. There are 19 longitudinal studies which form the CLOSER partnership and those which related to early life health are included in Table 1. Additionally, the UK Data Service and the British Library are included in the Consortium. CLOSER works closely with their partners to document how successful linkages can be achieved.	(O'Neill, 2019)
ScottisH Informatics Programme (SHIP)		SHIP is a publicly funded system that links data and makes it available for research whilst preserving privacy. SHIP has been able to integrate electronic patient records and non-medical routinely collected data, with the aim of providing a platform for Scottish record linkage. It was able to link data from Scottish morbidity records, Stillbirth and Infant Death Survey and Birth Certificate Database of live births in Scotland to investigate the effect of time and day of birth on the risk of neonatal death at term. They found that delivering an infant outside the normal working week was associated with an increased risk of neonatal death at term.	(SILC, 2021; Scottish Informatics Programme, 2011).
Administrative Data Research Centre (ADRC) operating in Scotland, England, Wales		ADRC is a partnership between universities, the Economics and Social Research Council, national statistical authorities, government departments and researchers, working across a range of sectors such as health, housing, and education. ADRC has been pioneering research exploring the links between education and developmental outcomes and exploring the variation in child development based on socioeconomic status, birth weight and gestational age.	(Data Linkage Scotland, 2016)

Service	Linked data	Summary	References
and Northern Ireland.			
eData Research and Innovation Service (eDRIS)		eDRIS is a national safe haven supporting researchers by advising on which datasets are available, the processes to gain access and the strengths and weaknesses of the data.	(NIHR Applied Research Collaboration South London, 2021)
Administrative Data Research (ADR) UK	Government data to public service data.	ADR is a partnership made up of ADR England, ADR Scotland, ARD Wales, ADR Northern Ireland, and ONS statistics. They ensure that data provided by UK government bodies is safely and securely accessed by approved researchers. There are several data linkage projects being undertaken across the partnership within the theme of children and young people. Notably, ADR is funding the generation of the ECHILD database.	(ADR UK, 2021c)
Centre for Longitudinal Study Information and User Support (CeLSIUS)		CeLSIUS funded by the is funded by the Economics and Social Research Council and provides support for users of the ONS LS. The CeLSIUS database is also where the research and output from the ONS LS can be found.	(University College London, 2021)
The Census & Administrative data LongitudinaL Studies Hub (CALLS-HUB)		CALLS-HUB was commissioned by the Economics and Social Research Council to support, promote and harmonise the work of the three Longitudinal Study Research Support Units ONS LS, SLS and NILS.	(CALLS-HUB, 2021a)
Abbreviations: Hospital Episode Statistics, HES; General Practitioner, GP; Office for National Statistics, ONS; National Health Service, NHS; Medical Research Council, MRC; National Pupil Database, NPD; Medicines and Healthcare products Regulatory Agency, MHRA; National Institute for Health and Care Research, NIHR			

Thus, the use of linked routine data for health research in the UK has increased over the last decade. Coupled with growing interest and investment by policymakers in data linkage, this creates the opportunity for better informed decisions around early years health services planning (Padmanabhan, *et al.,* 2019). Throughout a child's journey, from pregnancy to age five, they will likely engage with multiple services, so linking these data can create a clearer picture of that families' experiences and needs, to improve support provided by local services.

1.4 Health care decision-making in the UK

This section describes the key aspects of health care decision-making in the UK that are relevant to the research conducted in this thesis. This thesis focuses on decision-making made within policy and practice, that informs the provision of early life health services rather than the decision-making of individuals (i.e., members of the public's decision-making). As such, I will describe organisations that have the potential to benefit from the increased use of linked routine data.

1.4.1 Local Authorities

In 2012, the Health and Social Care Act introduced a major reform affecting the way public health is organised, commissioned, and delivered (Riches *et al.,* 2015). It involved key responsibilities regarding the protection and improvement of public health being transferred from the NHS to local authorities. It also created the new executive agency, Public Health England (now replaced by UK Health Security Agency and Office for Health Improvement Disparities), to deliver services at the national level.

A local authority is an organisation that is officially responsible for all the public services and facilities in a particular geographical area (Collins Dictionary, 2023b). They generally take responsibility for housing and transportation services, providing schools within their community, and social care services (Health and Safety Executive, 2023). As the health of the public is also influenced by wider determinants (such as housing, economic development, and transport), decisions made by local authorities can have profound impacts on these factors (Department of Health, 2012). Henceforth, I refer to decision-makers at this level as local decision-makers.

In October 2022, the NIHR announced a £50 million investment to boost local authorities' capacity and capability to conduct research (NIHR, 2022). They established 10 pioneering Health Determinants Research Collaborations (HDRCs) with the aim of embedding evidence-based decision-making within local authorities. The HDRC model

hopes to support a research culture within local authorities, building up experience of research as well as skills and expertise of academics. The intention is to build a partnership between local authorities and higher education institutions to help address the wider determinants of health. One focus is on co-producing research with the public, policymakers, and practitioners, to ensure research addresses the most important questions to services at the local and national levels (NIHR 2023). HDRCs promote a culture where everybody is responsible for research. Hence, HDRCs could benefit the use of linked data research by local authority decision-makers, as they can enable those decision-makers to develop the skills that allow them to effectively engage with research.

1.4.2 Integrated Care Systems

When the NHS was first established, the focus was on treating single conditions or illnesses. Hence, the NHS comprises of many organisations that operate independently from one another. However, our health and care needs have since changed. People are living longer and with multiple conditions. The UK government's National Data Strategy launched in 2020, highlights how public services are increasingly interconnected and that they need to work together to deliver for their populations (Department for Digital, Culture, Media & Sport & Department for Science, Innovation & Technology, 2020). Therefore, NHS organisations and local public services have been seeking to partner with each other, to better meet the needs of the population.

At the local level, GPs have joined together to form Primary Care Networks, where they work in collaboration across areas called *'neighbourhoods'*. This has allowed them to provide a broader range of services than what they could provide as a single general practice. Health care organisations have also been working together across larger areas known as *places*, which often cover the same area as a local authority (The King's Fund, 2022a).

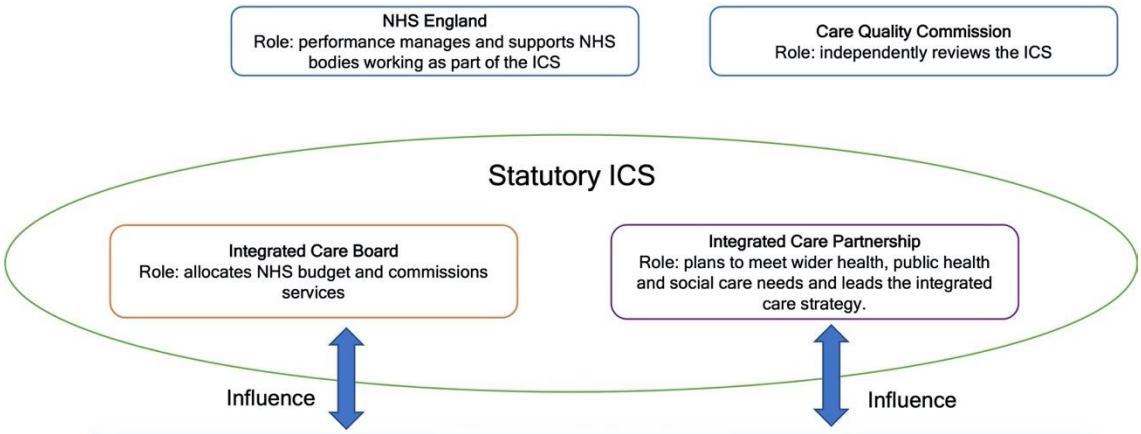
In February 2021, the Department for Health and Social Care published the White Paper *'Integration and innovation: working together to improve health and social care for all'*, which exemplified this move towards a model of collaboration, partnership, and integration in the health care system. It put in place a legal framework to enable services to work closely together with the aim of ensuring the health care system can better meet our changing needs (The King's Fund, 2022a; Department of Health and Social Care, 2021a).

Consequently, Integrated Care Systems (ICSs) were established in England in July 2022 (The King's Fund, 2022a). These ICSs are geographically based partnerships that bring

together organisations that provide and commission health care services. ICSs are made up of two parts: Integrated Care Boards (ICBs) and Integrated Health and Care Partnerships. ICBs are statutory organisations that decide how the NHS budget for their area is spent. Integrated Health and Care Partnerships bring the NHS together with other key partners, like local authorities, to develop a strategy to enable the ICS to improve health and wellbeing in its area.

ICSs are mainly funded by NHS England, which is the national body for the NHS in England and sets the operational priorities for the health system. The Department for Health and Social Care sets out what the NHS is expected to deliver as a result of the money it receives from the government. Figure 2 summarises the structure of an ICS, adapted from The King's Fund (2022b).

Figure 2 Summary of the ICS structure*



Partnership and delivery structures		
Name	**Participating organisations**	**Geographical area**
Provider Collaboratives	NHS trusts and, as appropriate, voluntary, community and social enterprise (VCSE) organisations and the independent sector.	System level: usually covers s population of 1-2 million.
Health and Wellbeing Boards	ICS, Health Watch, local authorities and wider membership, where appropriate.	Place: usually covers a population of 250-500,000.
Place-based Partnerships	Can include ICB members, local authorities, VCSE organization, NHS trusts, Health Watch and primary care.	Place: usually covers a population of 250-500,000.
Primary Care Networks	General Practice, community pharmacy, dentistry, opticians.	Neighbourhood: usually covers a population of 30-50,000.

**Figure adapted from 'Integrated care systems: how will they work under the Health and Care Act?' by The King's Fund (2022b) used under a CC BY-NC 4.0 licence.*

An example of an ICS is the West Yorkshire Health and Care Partnership (WYHCP). This ICS supports 2.4 million people and 22% of those people are living in areas ranked in the most deprived 10% in England. In November 2022, the WYHCP had 291 GP practices, 547 community pharmacies, 277 dentists. 431 at home service providers and 52 Primary Care Networks (West Yorkshire Health and Care Partnership, 2023). They work in partnership with NHS organisations, local authorities, charities, and the community voluntary and social enterprise sector to improve the health and wellbeing of

people living in their five local places: Bradford District and Craven, Calderdale, Kirklees, Leeds, and Wakefield. The WYHCP emphasise the importance of communicating with their partners, stakeholders, and members of the public to plan and design their services to ensure it meets the needs of the community. Their agreed priority areas include improving outcomes for maternity, children, and families (West Yorkshire Health and Care Partnership, 2023).

As many of the decisions impacting public health, including those that impact the early years of life, are made at the local level, this thesis considers how linked routine data can be used to inform these decisions.

1.4.3 Data driven decision-making

The UK government has stated a commitment to using data to improve the health and care of the population, in a safe, trusted, and transparent way. These intentions were documented in the *'Data saves lives: reshaping health and social care with data'* policy paper (Department of Health and Social Care, 2022). They describe the goal of having a health and care system that is underpinned by high quality, readily available data. As part of this strategy, they aim to build analytical and data science capability, and collaborate with wider partners to support local decision-makers to use data.

The concept of a *'learning health system'* describes how data from the health system can be used to generate new knowledge, which can then inform how health and care services are delivered. It is the idea that the health system learns from every patient interaction, to continually improve services and better understand the health and care needs of their population. As such, routine data collected by the NHS is an important aspect of a learning health system.

A learning health system involves: 1) collecting, (2) assembling and (3) analysing the data, (4) interpreting the results, (5) representing, (6) managing and (7) applying the knowledge, and then (8) changing practice as a result of this. It is depicted as a cycle, shown in Figure 3, which is adapted from Flynn *et al.*, (2018). Hence, linking routine data at a local level can enable local learning health systems.

Figure 3 Learning Health System Cycle*

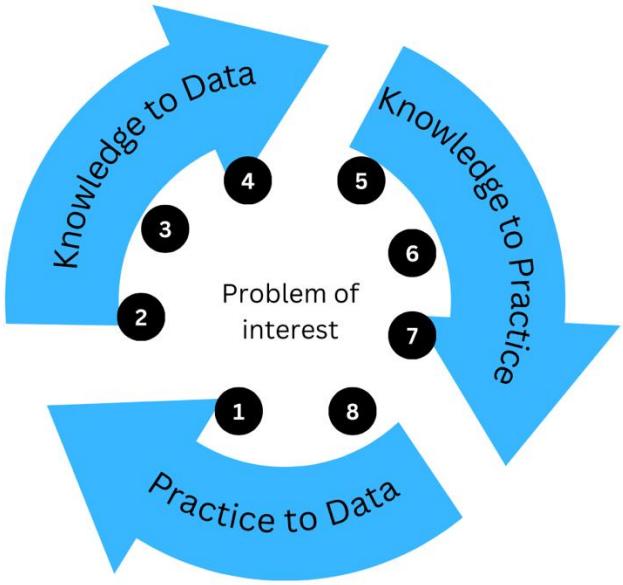


Figure adapted from 'The learning health cycle of the learning health system with 3 information flows and 8 steps' by Flynn, et al., (2018) used under a CC BY-NC 4.0 licence.

Different types of evidence may be needed for health care decision-making. For example, *'evidence of need'* might be required, which provides a picture of the issues faced in a community. This ensures that the most appropriate services are implemented and match the needs of that local population. Therefore, data collected by health, education, or social services could be useful. Once an issue has been identified, evidence of *'what works'* might be needed to address the issue effectively. Finally, *'evidence of impact'* can help decision-makers understand whether the programme has made a difference to the intended beneficiaries. It may help justify difficult decisions about which services should be funded as well as identify practical issues with the service delivery so that changes can be made, and stronger outcomes achieved (The National Lottery Community Fund, 2022).

Often, it is important for decision-makers to consider evidence that is related to their local context. An intervention with a strong body of evidence in Australia, may not have the same impact on service uses in the UK. Hence, a combination of types and sources of evidence is often necessary to build a more complete picture of an issue. It often involves finding the point where research, stakeholder opinion, practitioner experience, and local context align (The National Lottery Community Fund, 2022).

The United States Department of Health and Human Services produced a guide to data-driven decision-making, where the process is comprised of four stages, illustrated in Figure 4 (James Bell Associates, 2018).

Figure 4 Stages of the data-driven decision-making process*

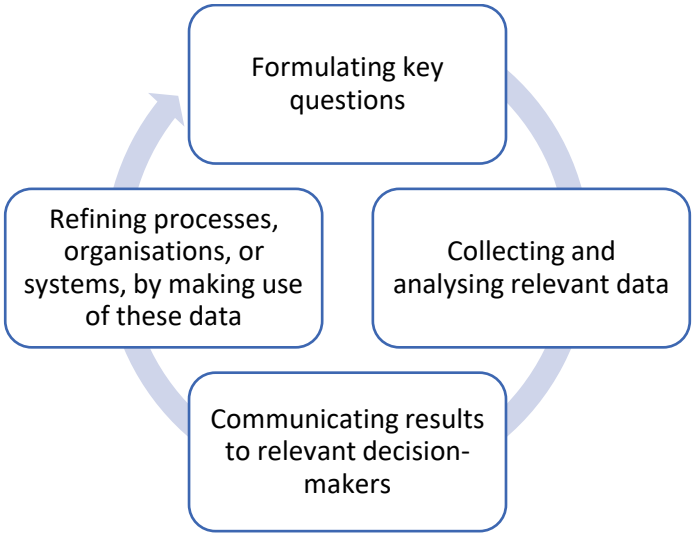


**Figure adapted from Stages of the Data-Driven Decision-Making Process by James Bell Associates (2018).*

The idea of using data to drive local decision-making underpins the theoretical model presented in Chapter 3, which informs the research conducted as part of this thesis. The model set out in Chapter 3 focuses on how a specific linked data research cohort can be used to inform local decision-making around early life health.

1.5 Research utilisation

Increasingly, governments are investing in research to help create an evidence base for better decision-making, including those discussed in section 1.3.1. The term *'Evidence Based Policymaking'* (EBP) has been used to describe the need for policy decisions to be informed from *"rigorous and accurate use of scientific evidence"* (Parkhurst, 2017 pg.20). The term '*data-informed decision-making'* is also used and refers to the use of available evidence, from a range of sources, to inform policymaking (The National Lottery Community Fund, 2022). In 2013, a network of *'What Works Centres'* were launched to support the use of evidence in UK policymaking, where each centre supported the dissemination of research on a range of topics. These included education, health, social care, and early intervention (The National Lottery Community Fund, 2022).

Despite the movement for policy to be informed by scientific research, a gap remains between scientific research findings and policy and practice. EBP is often used

interchangeably with related terms such as *'knowledge translation', 'knowledge action', 'knowledge exchange', 'knowledge mobilisation'* and *'research translation'.* Each of these terms acknowledges the gap between research and policy and the importance of closing this gap to allow for better informed decision-making.

In health care, this gap between research and policy can contribute to health inequalities and can lead to under-use of effective treatments and over-use of unhelpful treatments (Ward *et al.,* 2009). As this is negatively and unequally impacting health, this has led to increased emphasis on transferring knowledge into action.

In this section of the chapter, I will utilise the work of Katherine Smith, Katherine Oliver, Annette Boez, and Paul Cairney, to set out the key theories around research utilisation by policymakers.

1.5.1 Research-policy gap

Extensive literature discusses the gap between research and policy whereby there is a disparity between those producing the evidence and those involved in constructing policies (Smith, 2013). This is in response to the assumption that research evidence is being routinely ignored by decision-makers (Ham *et al.,* 1995; Black and Donald, 2001; Smith 2013). The field of EBP offers some practical insights into the factors influencing research use and the strategies to improve the use of research by policymakers (Innvaer et al., 2002; Orton et al., 2011; Oliver et al., 2014).

Oliver *et al.,* (2014) conducted a systematic review investigating the barriers and facilitators of evidence use by decision-makers, expanding and updating a previous systematic review (Innvaer et al., 2002). The majority of the papers identified in this review related to health research use. A summary of the barriers and facilitators affecting the use of research by policymakers can be found in Table 3.

Table 3 A summary of barriers and facilitators affecting the use of research by policymakers*

Barriers	Facilitators	Both a barrier and a facilitator
Timing and opportunity to use the evidence	Contact, collaboration and relationships between policymakers and research staff	Availability and access to research
Policy maker research skills	Relationship with policy makers	Clarity, relevance, and reliability of research findings
Policy maker beliefs of the utility of evidence	Relationship with researchers	
Organisational resource constraints, such as costs	Researcher understanding of the policy process and the context surrounding policy priorities	
Lack of managerial support and staff turnover		
Poor long-term policy planning		

**Table adapted from Oliver et al., (2014)*

The most frequently reported barriers to research utilisation included poor access to high quality, timely, and relevant research. In terms of facilitators, collaboration between researchers and decision-makers, improved relationships, and skills-building were frequently reported. Oliver *et al.,* (2014) found few studies that provided empirical data regarding policy processes, making it challenging to describe the role of evidence and other factors influencing policy. They also identified researcher' understanding of the policy process and the context surrounding policy priorities as key for research translation into policy, and that research evidence is just one source of information for policymakers. Thus, by understanding the perspectives of the decision-makers towards the policy process and the use of evidence, researchers can better understand how to influence policy. Finally, Oliver, *et al*., (2014) explained how the expansion in policymakers within the UK healthcare setting has broadened the range of potential *'evidence-users'*, making it timely to investigate issues surrounding research evidence uptake. Since this review was published, the move towards an ICS model has further increased the number of potential evidence users. These findings motivated the aims and objectives of the research detailed in this thesis.

Smith (2013) presents the recommendations based on five literature reviews (Contandriopoulos *et al.* 2010; Innvær *et al.* 2002; Mitton *et al.* 2007; Nutley, *et al.* 2007; Walter, *et al.* 2005) regarding how to increase the use of research in policy and practice. All five reviews recommended ensuring the research was accessible, such as providing clear and concise summaries personalised to the research user. It was also recommended that ongoing and collaborative relationships are developed between researchers and policymakers, as this can build trust and allow for discussions around

the importance of a particular policy issue or how potential solutions can be assessed. This aligns with the findings of Oliver, *et al*. (2014). Authors of the five reviews also suggested that improving structural communication channels would improve evidence use, for example, by investing in knowledge-transfer training or in *'knowledge brokers'.* Knowledge brokers facilitate knowledge translations by providing this link between researchers and research users (Smith, 2013). Ensuring there are sufficiently high incentives for researchers and research users to engage in knowledge exchange was also suggested. An example of this includes reducing the *'costs'* associated with engaging in knowledge-transfer activities.

1.5.2 Theories of knowledge transfer

During the 1960s and 1970s, before the term EBP was adopted, policymakers made efforts to improve research use. These efforts stimulated the generation of theories, in the 1970s and early 1980s, regarding the relationship between research and policy (Smith, 2013). A variety of theories or *'models'* are summarised in Table 4.

Table 4 A summary of the models regarding the relationship between knowledge and policy*

	Model	Explanation	Key authors
1	A knowledge-driven model	Research findings or knowledge provides the necessary pressure for policy to develop in line with this knowledge.	(Davies, Nutley, and Smith, 2000; Weiss, 1979)
2	A problem-solving model	A policy problem is first recognised, which stimulates research that aims to provide the evidence base for policy solutions.	(Davies, Nutley, and Smith, 2000; Weiss, 1979)
3	A political model	Research is used in a pre-specified manner to support policies that have already been considered for implementation due to political reasons. This highlights the dominance of political values within the policy process. Thus, research will only play a role if it is consistent with the dominant ideologies.	(Weiss, 1979)
4	A tactical model	Policymakers encourage or fund research to delay the decision-making process when facing awkward decisions or to distract attention. E.g., the UK government commissioned a review of health inequalities in 2010, when it was clear they were not on track to meet their inequalities target (Marmot et al., 2010).	(Weiss, 1979)
5	A two-community model	Portrays policymakers and academics as contrasting communities with different and often conflicting values, reward systems and languages. Thus, making it difficult for academic research to inform policy in a meaningful manner.	Caplan (1979)

	Model	Explanation	Key authors
6	An interactive model	Research is just one factor of many that can influence policy. Other important factors include political ideology, external pressures and the personal experiences of the policymakers involved. It is similar to the 'garbage can' model by Cohen et al., (1972) which suggests an interactive process whereby a range of actors and interests feed into it and outcomes are challenging to predict. This built-in complexity means that chance can play a significant role.	Donnison (1972) Weiss (1979)
7	An enlightenment model	Research influences policy indirectly and in a diffuse manner by gradually changing the way policy actors think about policy issues over long periods of time. It can lead to a change in the way policy problems are framed and conceptualised, rather than addressing specific problems. This model suggests that it is the ideas associated with a body of research that influences policy rather than individual research studies. This model remains popular despite being developed in the 1970s. However, it offers little guidance for those seeking to improve the relationship between research and policy. It is also considered to be time-consuming as it requires changing the perception of research. Weiss (1977, 1979, 1982) argues that there is little potential for research to directly impact on policy outcomes.	Weiss (1977, 1979, 1982)

* *Table summarises the models presented by Smith (2013).*

Models 1 and 2 describe a direct, linear connection where policy-relevant research findings drive change (knowledge-driven model) or provide direct solutions to problems already identified by policymakers (problem-solving model). For many public health researchers, tobacco-control policies in the UK represent the positive influence of public health research on policy and is an example of the knowledge-driven model. As such, the knowledge derived from scientific research helped identify a problem, the problem being tobacco causes harm to health, which subsequently informed key policy response. It could also be understood through the problem-solving model, where researchers focus their attention on assessing policy interventions that are most likely to be effective in addressing this tobacco related harm (Smith, 2013). The linear relationship between research evidence and policy is implied throughout the field of EBP and has become increasingly embedded in UK policy and higher education. If this linear model is accepted, then the value of research is based on its impact on policy (Black, 2001). However, these models have been criticised as they fail to capture the complexities of the relationship (Smith, 2013; Black 2001).

The five remaining models provide alternative conceptualisations, which are united by the view that policymakers rarely utilise research in this direct and linear way. The political model of the relationship between research and policy has been shown to be apparent in the UK, where research that challenges political values has been ignored, notably in the case of illegal drugs. The scientist David Nutt was sacked from the Advisory Council on the Misuse of Drugs for a publication in The Guardian newspaper challenging the government's decision to reclassify cannabis into a more harmful category (Smith, 2013). However, critics have argued that the political model is still simplistic for many policy issues (Davies *et al.,* 2000).

Whilst the *'two-communities model'* is not always directly referenced, existing assessments of the use of research in health policy and practice highlight the perceived *'gaps'* between researchers and decision-makers. For example, Walt (1994) and Saunders (2005) argue that lack of understanding and interactions between policymakers and researchers is a key barrier to research utilisation. Lomas (2000) and Lavis (2006) highlight the importance of achieving a shared understanding between the researchers and policymakers for research utilisation. These authors tend to assume that research utilisation would be improved if policymakers could better access and understand the research, which can be achieved by improving mechanisms of communication and levels of trust between the two communities. However, the problem with this, is that policymakers and researchers often have fundamentally different ideologies, which can result in disagreements about what the research findings imply for policy. This was demonstrated by Bartley (1988,1992) who investigated how public health messages were constructed and translated into policy. Bartley (1988,1992) highlighted that the disconnect between researchers and policymakers may be the result of their contrasting disciplinary training, where those with the same disciplinary training find it easier to communicate with each other. The *'two-communities model'* has also been criticised for excluding other important actors who are influential in the knowledge transfer process such as journalists (Lindquist, 1990).

The *'enlightenment model'*, proposed by Weiss in the 1970s, views research as one of several sources of knowledge. This model was extended during the 1980s and 1990s to a more interactive model, which is based on a close dialogue between researchers and policymakers. More recent developments of this theory describe the *'knowledge creep'*. This implies that research influences policy over time through gradual changes to actors' perceptions and ways of thinking, rather than the results of direct impacts. Accepting this model suggests that policymakers need to benefit in some way if they are to use it. Hence, researchers need to take the full complexity of the policy situation into account if their research is to be utilised. Researchers must also recognise that there are other

influences on policy (social, political, ethical, cultural, economic) which must be accommodated and that the most likely way to influence policy is through extended communication (Black 2001). These considerations are important when exploring how linked data research can inform decision-making.

Boswell and Smith (2017) highlight that whilst there is extensive literature on the *'research-policy'* relationship, there are fewer contributions regarding guidance on how to achieve research impact. Existing guidance and models are underpinned by linear ideas about the relationship between research and policy. Boswell and Smith (2017) draw on the wider social science literature to theorise the relationship between research and policy, relating this to impact. They propose four different approaches: (1) knowledge shapes policy; (2) politics shapes knowledge; (3) co-production; and (4) autonomous spheres. These theories are summarised in Box 4, where each theory has implications for how we define impact.

Box 4 The research-policy relationship theorised by Boswell and Smith (2017)

Knowledge shapes policy

This approach focuses on the perceived gap between research and policy communities, where research can be relevant to policy, but communication problems hinder its impact. As such, research may not be presented or disseminated in a way that is relevant or accessible to policymakers or they are unable to action the research due to resource constraints. Boswell and Smith (2017) suggest that, under this approach, the flow of knowledge from research to policy can be improved by taking practical steps. Practical recommendations in the literature can be found in section 1.5.1.

Politics shape knowledge

A notable critique of the *'knowledge shapes policy'* model is that policies can also shape research and hence the use of research. Hence, the *'politics shapes knowledge model'* implies that research will only be used by policymakers if it supports the dominant political ideologies.

Co-production

Research knowledge is co-produced through ongoing interactions between researchers, policymakers, and a range of other actors including politicians, journalists, and lobbyists. Hence, in trying to measure impact, it would be challenging to identify the subtle and incremental process through which a range of actors influenced ideas, but also how the ongoing feedback of these ideas shifted the behaviour of these actors that gradually changed political behaviour.

Autonomous spheres

This radical view suggests there is no overall causality between research and policy, where instead the two systems selectively pick up on signals from the other system. A version of this account is Caplan's (1979) *'two communities'* theory which identifies a cultural gap between policymakers and researchers. The challenge is to understand how each system selectively picks up these signals. Thus, we need to understand how the political system makes sense of its environment and draws selectively information being signalled from scientific research. *'The garbage can'* model developed by Cohen *et al.,* (1972) offers a way of theorising how different ideas are picked up, depending on the political problem. This approach is cautious in attempting to demonstrate impact and instead suggests focusing on how and why a political system picks up evidence from social science.

Theories of *'institutionalism'* are often used to explain the impact of institutions and organisations on policy and research. This suggests that policy processes are shaped

by historically constructed institutions, where ideas have become embedded, rather than the collective result of individuals. From this perspective, it becomes increasingly difficult to change the direction of policy as previous decisions have become embedded in the way of thinking. Smith (2013) describes how *'institutionalised ideas'* around health and economics make it more likely that research-based ideas that support these *'institutionalised ideas'* will move into policy, and those that challenge these ideas will not. Hence, bringing researchers and policymakers together will likely exacerbate the effects of policy on research.

This leads to concerns over academic independence when working closely with policymakers. It can dampen efforts to promote radical ideas that do not align with government policies. Cairney and Oliver (2018) suggest that policymakers and *'influencers'* who regularly engage with each other begin to conform with each other's beliefs which may reduce researchers' ability to be independent and objective. Thus, theories of institutionalism offer an alternative perspective to the two-communities model, where it is assumed that collaboration between research and policy is beneficial.

Based on the theories of knowledge transfer I have presented, the use of research in decision-making is unlikely to be a one-way process. Research can both inform and be informed by policy. Rein (1980) argues that it is more useful to think of the *'interplay'* between research and policy, highlighting the interactive relationship between them.

1.5.2.1 Learning from political science theories

Smith (2013) suggests that we can glean some important insights from political science around the theories of policy change. A summary of theories from political science that can help us understand knowledge translation is presented in Table 5 .

Table 5 A summary of theories from political sciences relevant to research translation

	Theory	Insights from the theory relevant to research translation
1	Resistance to change theory	This suggests that if policy is resistant to change, then it will be resistant to research that challenges it. These theories can be divided into two groups. The first group involves approaches that consider politics and ideologies as pivotal factors shaping policymaker decisions, where policymakers find it challenging to make changes that oppose the political, economic, and social context in which they sit. This implies that research may only be referred to when it supports the policy response. If we can understand how political ideologies are driving policy decisions, it could allow researchers to challenge these underlying interests. This theory has been criticised for downplaying the agency of individuals and non-dominant groups (Smith, 2013). The second group highlights the importance of historical decisions, whereby previous decisions often limit future decisions. As part of this theory, Immergut (1998) argues that the context in which the decisions are taken must be considered to understand why it was taken.
2	Incremental policy change	It is suggested that policymakers *'muddle through'* policymaking, considering a small range of options they judge to be feasible and implementing the option with the greatest consensus. Hence, policy change is gradual, and policymakers are capable of learning (Heclo 1974; Lindblom 1959). This group of theories considers *'policy networks'* which explore how knowledge is transferred between actors (including policymakers, academics, and other interested parties such as journalists or lobbyists) (Marsh and Rhodes 1992; Marsh and Smith 2000). Actors involved in the network are linked by a shared culture or interests. However, the shared culture between the actors, which hold together the network may limit the opportunity for radical new ideas or alternative ways of thinking. These networks are presented as constantly changing structures, shaped by the actors within them. Therefore, research informed ideas can enter the network, although change is unlikely to be radical as consensus among the actors is what allows these networks to exist (Smith, 2013).
3	Theory-informed policy studies	The concept of bounded rationality, which describes the limits of our cognition capacity, can be helpful in understanding the constraints on human decision-making (Cairney and Kwiatkowski, 2017). Policymakers lack time and resources to consider all possible information, in addition to all the consequences of their actions, when making decisions. Hence, they develop heuristics to allow them to make good choices (Cairney and Oliver, 2018). Acknowledging these cognitive biases and not holding policymakers to an information processing standard that no human possesses, can be helpful when designing strategies to promote the use of research findings. Typically, policymaking is depicted as a cycle where there is a series of well-defined stages (agenda setting, policy formulation, legitimation, implementation, evaluation and policy maintenance, succession, or termination), where researchers know when and how to present evidence (Cairney, 2013). However, policy-theories also acknowledge the involvement of many actors who influence decision-making, and that researchers compete with other actors to present their evidence.

A common theme throughout the political theories presented is the importance of the many actors involved in promoting research, as well as the complexity of the policymaking process. The importance of the different actors in shaping decision-making is consistent with the *'interactive model'* of knowledge transfer.

Many of these theories suggest that significant policy changes are hard to achieve, where research findings challenge existing political ideology. This aligns with theories of *'institutionalism'* and the ideas of Boswell and Smith (2017) in the *'politics shapes knowledge'* model and the *'political model'* of knowledge transfer, presented in section 1.5.2. The implication of this is that researchers are unlikely to conduct research that challenges existing policy. Smith (2013) discusses how the stakeholders in policy and research spaces believe that funding bodies, including research councils are motivated by political and policy agendas. This belief results in researchers framing their funding applications accordingly, which minimises the likelihood of researchers applying for funding to explore research ideas that are perceived to challenge current policy. This also suggests that researchers working with ideas that complement those within the existing policy frame, will find it easier to influence policy. If this is the case, this could have implications for how linked routine data are used for research.

The theories around incremental policy change align with the *'enlightenment model'* proposed by Weiss (1977, 1979, 1982), which implies that research can influence policy over time by gradually changing actors' perceptions. Hence, Cairney and Oliver (2018) suggest that researchers invest their time in building longer term alliances with policymakers, to gain knowledge of the political system. This will allow researchers to take advantage of opportunities for policy change and gradually influence the ideology of policymakers. However, as discussed in section 1.5.2., building these alliances between researchers and decision-makers can reduce a researcher's ability to conduct independent and objective research, as their beliefs begin to conform with those of policymakers.

1.5.2.2 The power of ideas over evidence

Weiss (1982) argues that the ideas associated with a body of research are what influences policy rather than individual research studies. As such, the ideas from multiple studies indirectly influence policy through articles in the academic literature, the media, through advice of lobbyists, and conversations with colleagues (Weiss, 1982). Focusing on the use of ideas rather than specific evidence can help overcome some of the limitations associated with the existing theoretical frameworks described in section 1.5.2.

The notion that ideas impact policy, rather than evidence, is complex because ideas are constantly changing and being reconsidered as actors communicate with one another. This makes it challenging to trace the influence of ideas on policymaking. In addition, actors involved in a policy decision may not be aware of the ideas that have shaped their decisions. Smith (2013) suggests that it is more accurate to perceive the interplay between research and policy as a continual exchange and translation of ideas rather than a process of research utilisation. This idea is supported by Haynes *et al.,* (2011) who looked at public health utilisation in Australia and found that politicians tended to view researchers as a subgroup of experts and one of several sources of ideas, not necessarily evidence. If this idea is accepted, then the findings from linked data research can be a potential source of ideas which inform decision-making.

In addition, Smith (2013) introduces the notion of *'charismatic ideas'* to explain how researchers can achieve radical policy changes. A *'charismatic idea'* is one that can provide a convincing vison of an alternative future that policymakers can support. This draws on Max Weber's notion of *'charisma'*. Weber links charisma to individual *'leaders'* and Spencer (1973) suggests that the central feature of a *'charismatic leader'* is being able to convince others that their vision of the future will occur (Eisenstadt, 1968). If actors are persuaded of this vision, then there are more likely to make decisions based on this belief. Charismatic ideas emerge when people '*think outside the box'* of what is rationally accepted. Applying this notion to the relationship between public health research and policy, it is expected that charismatic ideas will be initially contested as they challenge the status quo, but eventually become accepted. An example of this was the way in which tobacco-related research was able to influence UK policy between 2000 and 2011. Tobacco advocates were able to convince policymakers that the tobacco industry was not a legitimate policy actor, and that tobacco is *'uniquely damaging'*, without challenging the accepted idea that the wider business community are legitimate policy actors (Smith, 2013).

1.5.3 Support for research impact in the UK

The rise of the New Labour government leading up to the 1997 UK general election renewed focus on EBP. This was documented in *'The Modernising Government White Paper'* where the government committed to using research to better understand the problems they were trying to address (Cabinet Office, 1999). They stated that all policies should be appropriately evaluated so that lessons can be learnt from success and failures. However, in using the term EBP, they often referred to the linear relationship between research and policy as seen in models 1 and 2 in section 1.5.2.

Academia's interest in research utilisation is backed by the Research Excellence Framework (REF), which assesses the performance of UK universities. Since 2014, the REF has increased funding for those who are able to demonstrate research impact. In REF 2021, funding for impact increased to 25% and it will remain at that level for REF 2028 (Wilsdon *et al.,* 2023). Research impact is defined as the effect of research beyond academia where the benefits apply to one of more areas of *"the economy, society, culture, public policy or services, health, the environment or quality of life"* (UKRI, 2022a). UK policymakers and academics continue to support the production of policy-relevant research. This could lead researchers to focus on research that is likely to align with policymaker's existing ideas, as this is more likely to achieve research impact, which leads to increased funding.

Moreover, the REF is based largely on the linear model of research utilisation where policymakers utilise expertise from research to inform more effective policies. This places the emphasis on individual researchers to explain how their research has achieved impact. Critics of the REF argue that there should be a shift from individual researchers to supporting collective diffusion of larger bodies of research, which is aligned with Weiss' *'enlightenment model'*. This could be achieved by supporting researchers from across a diverse range of studies on a particular policy topic to work together (Boswell and Smith, 2017). Nonetheless, challenges remain in tracing impact, as ideas are poorly monitored over time.

The *'Rethinking Policy Impact Project'* explored how to rethink the UK's approach to policy impact and recommended that frameworks supporting policy impact should be guided by six core principles (Boswell *et al.,* 2022):

1. **Collaboration:** Researchers should be incentivised to work collaboratively, rather than seeking sole credit for the research of their individual team or institution.
2. **Bodies of knowledge:** Researchers should be encouraged to contribute to, and effectively communicate, a wider body of knowledge and insights.
3. **Equality and diversity:** There should be support for those with protected characteristics and at an early career stage to diversify the research informing policy.
4. **Quality of policy engagement:** Productive engagement should be rewarded in policy impact frameworks.
5. **Public and community engagement:** Researchers should be incentivised to contribute to informing and enriching the parameters of public debate on policies.

6. **Disruptive research:** Innovative, blue skies or disruptive research should not be devalued or crowded out by support for policy impact.

In 2005, the *'Evidence and Policy, a Journal of Research, Debate and Practice'* was launched, which was dedicated to the relationship between research and policy. However, the new body of literature did not appear to be building upon the work developed in the 1970s and 1980s, instead it focused on New Labour's conceptualisation, which depicted a rational relationship between research and policy decisions (Smith, 2013).

The UK's research impact agenda can be seen as a genuine attempt to incentivise and reward scholars for promoting the use of their research by policymakers and practitioners (Smith, 2013). The previous system placed emphasis on peer reviewed publications that largely appeal to an academic audience. There is a wealth of practical knowledge from researchers who have previously engaged in impact activity. However, the advice appears to be based on normative understanding of research and policy leaving new researchers to learn the same lessons through trial and error. Oliver and Boez (2019) suggest that more work is needed to understand how funders and universities can support researchers to do impactful research.

1.5.4 Mechanisms for achieving research impact

Despite the strong encouragement for researchers to create impact from their research, there are few empirical evaluations of the strategies aimed at achieving research impact. Oliver and Cairney (2019) explored *'how to'* advice for influencing policy and practice. They found limited empirical evidence on how academics can create impact with their research but found there was an abundance of advice based on personal experiences. This advice had good intentions but lacked understanding of the policy process. Oliver and Cairney (2019) have condensed the existing advice on how academics should engage with policy to achieve impact into eight main tips, detailed in Box 5.

Box 5 Oliver and Cairney (2019) eight main tips on how academics should engage with policy to achieve impact

1. **Do high quality research**

   Researchers are advised to conduct high-quality, timely and policy relevant research that is presented in an easy-to-understand format. The need for interdisciplinary research to provide new perspectives was apparent and that the practical significance of the findings should be explored.

2. **Communication of research**

   Researchers should communicate the research effectively, provide clear summaries of the problems and solutions and disseminate widely so that policymakers can follow up if they have questions. Established storytelling techniques can be used to communicate scientific ideas and engage the policy audience.

3. **Understand policy processes and policy context**

   Researchers are advised to learn about how policy works to help manage expectations about how research is likely to influence policy. It is also recommended that researchers take the time to listen and learn about their policy colleagues and understand what works best for them. It is important for academics to place their research within the policy context and not just within the context of academic literature. To do this, it involves learning what, where, when and who to influence and maximising ways to engage.

4. **Be accessible to policymakers**

   Building trust and developing relationships with policymakers is important. Building and maintaining meaningful relationships can be time-consuming and require a lot of commitment. The literature recommends developing leadership and communication skills to effectively develop these relationships. Several sources referred to the ideas of two-communities of policy and research which each have their own language and values. If researchers can learn to speak the policy language it can facilitate this meaningful relationship.

5. **Issue advocate vs honest broker**

   Researchers can simply disseminate their research honestly, clearly and in a timely manner, acting as an honest broker of the evidence. Actors may then use this research in a number of ways and researchers remain neutral in this process. Alternatively, researchers can recommend specific policy outcomes or describe the implications of the research and the preferred policy actions.

6. **Build relationships with policymakers**

   Building longer term networks create more opportunities for researchers to inform policy agendas, develop credibility and give better insights into policy problems. The best way to achieve the goals of getting research into policy is through co-designing research. These relationship building activities require significant investments and skills. It is cautioned that collaboration can often result in conflict. Hence, ground rules should be established to define how, when, and why to engage, recognising the effort that goes into it.

7. **Be entrepreneurial**

   Much of the advice portrays a persuasive researcher, that is comfortable in the policy environment and is always available when needed. Researchers are advised to develop media skills to effectively convince people that shared action is possible.

8. **Continuous reflection on engagement**

   Researchers should be genuinely motivated to take part in policy engagement, understanding the value of the exercise in its own right, rather than an activity that improves the stated impact of research. Recommendations are around learning and reflection on engagement actions to help train a new generation of impact-ready entrepreneurs.

These tips reflect some of the facilitators affecting research use by policymakers identified by Oliver *et al.,* (2014), including the importance of high quality, accessible research. The tips in Box 5 also reflect the more complex interaction between research and policy, emphasising the importance the relationships between researchers and policymakers. Tip seven, *'be entrepreneurial'*, resonates with Smith's (2013) *'charismatic idea'* and the need to be convincing. Hence, the tips presented in Box 5 can be considered when conducting research using linked routine with the aim of informing or contributing to policy.

Oliver et al., (2022) aimed to identify activities and initiatives that maximise the engagement of policymakers with research outputs (research-policy engagement), and the effectiveness of these approaches for research impact. They identified 428 organisations globally that have promoted research-policy engagement, where the majority are university based. They were able to identify nine types of research-policy engagement practice, summarised in Box 6. Although, these initiatives were not based on existing evidence or theory, and most were unevaluated.

Box 6 A summary of the types of research policy engagement practice identified by Oliver *et al.,* (2022)

Disseminating and communicating research

Formal institutional requests for evidence

This involves pulling evidence and expertise using formal institutions such as science advisory committees. Evaluations suggest that a more diverse and appropriate evidence base could be achieved by putting more thought into the goal and purpose of the formal evidence request.

Facilitating access to research

Deliberate attempts to facilitate access to evidence include funder-led initiatives to promote partnership working and identify policy-relevant questions. Internally conducted evaluations suggest that these initiatives have the potential to support policy-responsive research.

Building decision-maker skills

This focused on training and capacity building. Training by policy intermediaries focused on understanding and using the evidence. Training was found to be too academic.

Building professional partnerships

A popular approach focusing on the creation of policy/practice-research collaborations. Factors facilitating success include linking related collaborations with funding or networking schemes (such as NIHR-funded Policy Research Units). The literature suggests that building long-term, mutualistic partnerships that promote collaborative working may be central to addressing barriers to improving evidence use.

Strategic leadership

An example of organisations advocating for evidence-informed decision-making or training and capacity building for individuals to develop strategic leadership includes the Royal Society of Edinburgh. They devote resources to combining academic expertise and assembling stakeholders to influence global policy discussions. In the UK, the establishment of many dedicated policy teams reflects attempts to embed policy skillsets and provide a strategy for knowledge exchange.

Rewarding impact

Prizes and awards for impact or best use of evidence. No evaluations available.

Creating infrastructure and posts

This includes Areas of Research Interest Fellowships to align the work of public research councils with departmental priorities. The creation of longer-term relationships to ensure sustainability of the project beyond the funded lifespan is also included.

The main activities were focused on improving research dissemination and communication or creating relationships. For many, this related to increasing the impact of one piece of research or pulling in evidence responding to a policy or practice need, which is only relevant if you accept the linear models described in section 1.5.2.

Oliver *et al.,* (2022) suggest that these types of approaches do little to address the fundamental barriers for policymakers to engage with research, such as practical, cultural, or institutional barriers, meaning they can demonstrate little impact on policy and practice. They suggested that the increased number of initiatives promoting the use of research could lead to competition between the initiatives, as policymakers are time constrained and are only able to engage with some activities. Therefore, without evaluating these initiatives, there is a risk of policymakers engaging in a less effective activity, while poor experiences of engagement could harm future attempts. Oliver *et al.,* (2022) conclude that despite this mass of activity, it fails to provide useful lessons for

those seeking to improve evidence use. They recommend identifying what different stakeholders want to contribute and get out of the research-policy engagement activities, to avoid wasting efforts on ineffective engagement.

An example of an organisation that works to maximise the impact of studies that link routinely collected data is CLOSER. CLOSER works closely with the government, policymakers, think tanks, and the third sector, to maximise the impact of the longitudinal studies it partners with, where many of these partners use linked routine data. Their strategy for achieving impact focuses on supporting continuous stakeholder engagement, particularly with policymakers. CLOSER submits evidence to government consultations on behalf of their studies to increase the influence of the research findings. They also broadcast policy alerts through a newsletter to update researchers on the latest policy developments and calls for evidence (CLOSER, 2018a). This strategy aligns with the idea that there is an interplay between research and policy, where research can inform policy, while policy can inform research. Furthermore, as part of the CLOSER Learning Hub, they provide resources and capacity building training aimed at those in academia, government, and the third sector to develop an understanding of the value of longitudinal research. These are important activities that could influence the use of linked data by policymakers.

1.5.5 Measuring research impact

In the literature, there appears to be a consensus that researchers are responsible for articulating the impact of their research beyond academia. This is underpinned by the assumption that impact can be measured and reported. It is also assumed that researchers own efforts to achieve impact play a significant role in explaining why some research achieves impact in non-academic contexts, whilst others do not (Boswell and Smith, 2017). Thus, researchers are required to provide evidence of how this impact was achieved. This implies there is a pathway to impact that can be traced through a series of concrete activities. These assumptions contradict the more complex models of knowledge transfer, such as the *'enlightenment model'*, which suggest that it is not individual research that influences policy but a body of research that contributes to a policy decision. Hence, there is a lack of consensus around ways of tracing research impacts.

There are three main approaches to assessing research impact: forward tracing approaches, backward tracing approaches, and evaluation of mechanisms to increase research use. Forward tracing approaches begin with the research outputs and trace forward into policy or practice settings (Molas-Gallart *et al.,* 2022). They rely on researchers' recollections of research use. Backward tracing approaches analyse the

policy and practice setting, where behaviour can be tracked back to the research (Smith, 2007). An evaluation of the activities used to increase research use, like those described in section 1.5.4, may only demonstrate immediate use of research, which creates challenges for assessing the impact over a longer time period (Morton, 2015).

Two common approaches for measuring research impact include the *'Payback Framework'* and the *'Research Contribution Framework'* for measuring the impact of research.

The Payback Framework, developed by Buxton and Hanney (1996), was found to be the most used approach in a narrative review of the literature for measuring research impact by Greenhalgh *et al.,* (2016). The Payback Framework consists of the seven stages of research ((0) conceptualisation, (1) inputs to research; (2) research process; (3) primary research outputs; (4) secondary research outputs such as policymaking; (5) adoption by practitioners and publics; (6) final outcomes) and five categories to classify the benefits or *'paybacks'* of the research. These include knowledge (e.g., academic publications), benefits to future research (e.g., training new researchers), benefits to policy, benefits to health and the health system, and broader economic benefits. This framework considers the different interactions between researchers, potential users of research, and reflects on the links and feedback loops connecting each stage of the research process, from research agenda setting to dissemination. Together, the categories in the framework help to capture the diverse ways that impact can arise (Greenhalgh *et al.,* 2016).

The Payback Framework encourages an assessment of the knowledge base to be conducted, at the point at which the research is commissioned. This helps determine whether the impact can be attributed to that research or whether impact occurred as a result of other research being conducted at the same time. It assesses impact through case studies, combining researcher interviews and documentary analysis to ascertain if claims of impact can be verified. This can be resource intensive and is, therefore, not accessible to all researchers. This assessment of impact is usually project-focused meaning it rarely explores the impact of a combination of activities that researchers might have employed to achieve this impact. This is evidenced by Meagher *et al.,* (2008) who argue that it is often easier to attribute impact to a researcher's full body of work rather than a particular project's findings. Thus, Greenhalgh *et al.,* (2016) suggest that project-focused impact models may be underestimating impact, which aligns with the '*enlightenment model'*.

The second approach which is widely used is the *'Research Impact Framework'*. This framework was originally developed for academics so that they could measure and monitor the impact of their own research. Hence, Greenhalgh *et al*. (2016) describes this

as a *'light touch'* approach. This framework is not designed to be used by third parties as a formal assessment of impact.

The reason that impact assessments continue to remain faithful to the simplistic model, despite criticism, is that they provide a reassuring narrative to both policymakers and researchers (Boswell and Smith, 2017). Policymakers can signal that their decisions are justified by referencing research, or use research as rationale for delaying a decision. Researchers can secure funding for presenting evidence about the impact of their research, as discussed in section 1.5.3. These accounts dismiss the complexity of the relationship between policy and research and imply a superficial relationship between research and policy, where the aim is to demonstrate research use per se, rather than the deeper aim of using evidence to make policy that improves outcomes.

GUS is a good example of a national cohort study that has demonstrated tangible ways in which their research has influenced health policy and practice. A summary of GUS can be found in Table 1. GUS produced a report in 2012 to demonstrate the impact of their research. It utilised the work by Sarah Morton on assessing the contribution of research and is based on an approach adapted from John Mayne's Contribution Analysis (Morton, 2015; Mayne, 2008). By using this approach, they acknowledge policymakers and practitioners are influenced by many factors and that research is only one of these factors. They appeal to the more realistic idea that research contributes to policy or practice change (Growing up in Scotland, 2012).

Much of their success in informing policy and practice was attributed to their engagement activities. GUS employ a dissemination officer to ensure their findings are widely accessible. Their findings are publicised through their website, social media, emails, newsletters, at annual conferences, and targeted briefings (Growing Up in Scotland, 2012). The report emphasised the importance of engaging with the right stakeholders and ensuring they are involved as early as possible in the study. However, it is worth noting that GUS is funded by the Scottish Government. Consequently, the Scottish Government are more likely to engage with the study findings compared with research they have not commissioned.

The impact of GUS findings was reported under three categories: immediate, intermediate, and final outcomes. In terms of immediate impact, they provide evidence suggesting that their engagement activities led to a greater awareness of the issues facing young children in Scotland. The evidence they provided includes statements from policymakers at GUS events and reference to presentations where GUS findings have been quoted (Growing Up in Scotland, 2012).

To assess the intermediate impact, the report reviewed how those stakeholders who engaged in the GUS study used the findings to initiate change. They provide examples of where GUS data have been cited in the development of policy. They also reveal that policymakers across the Scottish Government are involved in the questionnaire design and reporting of the findings which helps ensure the study is relevant to the needs of Scottish Policy (Growing Up in Scotland, 2012).

Finally, GUS acknowledges that whilst it is not feasible to measure the contribution of research in isolation from other factors, they suggest that the activities described have contributed to the children in Scotland having a better start in life with positive outcomes later in life. To monitor the contribution of research they looked at the wider changes through Scotland's National Performance Framework, Health and Education statistics and other surveys (Growing Up in Scotland, 2012).

This report clearly describes outcomes relating to the impact of GUS research on policy and practice and the strategies employed to facilitate this impact. The report suggests that engaging with decision-makers from the beginning, to raise awareness of the study and gain input on what is relevant for policy, increases the likelihood that those findings will be used. However, it is unclear from the report whether the research impacts were a result of the original cohort study interview data or a result of their data linkage research. This highlights the need for greater transparency regarding the impacts of linked data, which is important as this thesis focuses on how linked data research can be used by policymakers. The impact assessment was also conducted by GUS therefore, this is not an independent assessment of their impact on policy and practice.

1.5.6 Research utilisation frameworks

Building on the literature presented in this section, many frameworks have been developed to guide and measure the research utilisation process. However, there is lack of high-quality evidence about what works, in which settings, and with whom (Ward et al., 2009). These frameworks about research utilisation are untested, meaning their applicability and relevance is unknown. In this section, I present the Knowledge to Action Framework for Public Health and the PARiHS framework as these are widely referenced for translating health care research into policy and practice. I also present a third framework (FHI 360 Research Utilisation Framework) which builds on the ideas from the Knowledge to Action Model and the PARiHS framework.

Knowledge to Action Framework for Public Health

Wilson *et al.,* (2011) created the *'Knowledge to Action Framework for public health'* as a representation of the processes that can be used by researchers developing and testing health care interventions, or by practitioners gathering practice-based evidence. The framework consists of three phases: research phase, translational phase, and the institutionalisation phase.

The research phase includes developing and testing scientific advances to ascertain appropriateness for translation into policy and practice. The translational phase consists of the processes to ensure widespread implementation of the research. This includes, making the decision to translate the evidence, transforming the evidence into an actionable output, and developing appropriate structures to support the dissemination to evidence adopters such as policymakers.

PARiHS framework

The PARiHS framework (Promoting Action on Research Implementation in Health Services) represents successful implementation (SI) as a function (f) of the nature and type of evidence (E), the context in which the evidence is being introduced (C) and the way the process is facilitated (F) (Kitson, 2008).

$$SI = f(E, C, F)$$

The PARiHS framework incorporates key themes from the literature which include (Kitson, 2008):

- The need for strong evidence to justify implementation.
- Implementing research into practice is an organisational issue rather than individual issue.
- Implementation strategies need careful planning.
- Criteria for evaluating the impact of the intervention must be agreed before implementing the change.

Each of the factors included in the framework have sub-elements that can be rated (from low to high). A high rating on each factor means it is more likely to be successfully implemented.

FHI 360 Research Utilisation Framework

Kim et al., (2018) suggest that these frameworks are rarely applied due to the dynamic processes and the diversity of the actors involved. They also argue that there is limited support on how to implement them. Kim et al., (2018) adapted these two existing frameworks to develop the FHI 360 Research Utilisation Framework. It comprises of four stages (Foundational, Research, Translation, Institutionalisation) and three key types of actors: evidence producers, knowledge brokers and end users. This framework is presented in Figure 5 and Box 7 summarises the role of the actors in this framework. The FHI 360 Research Utilisation Framework is depicted as a cycle to represent the dynamic process where actors can stop and return to prior phase at any point. Each of the phases can be implemented concurrently, simultaneously, or in reverse. This allows for constant learning and evaluation between the phases, showing how policy and practice feeds back into research.

Figure 5 FHI 360 Research Utilisation Framework*

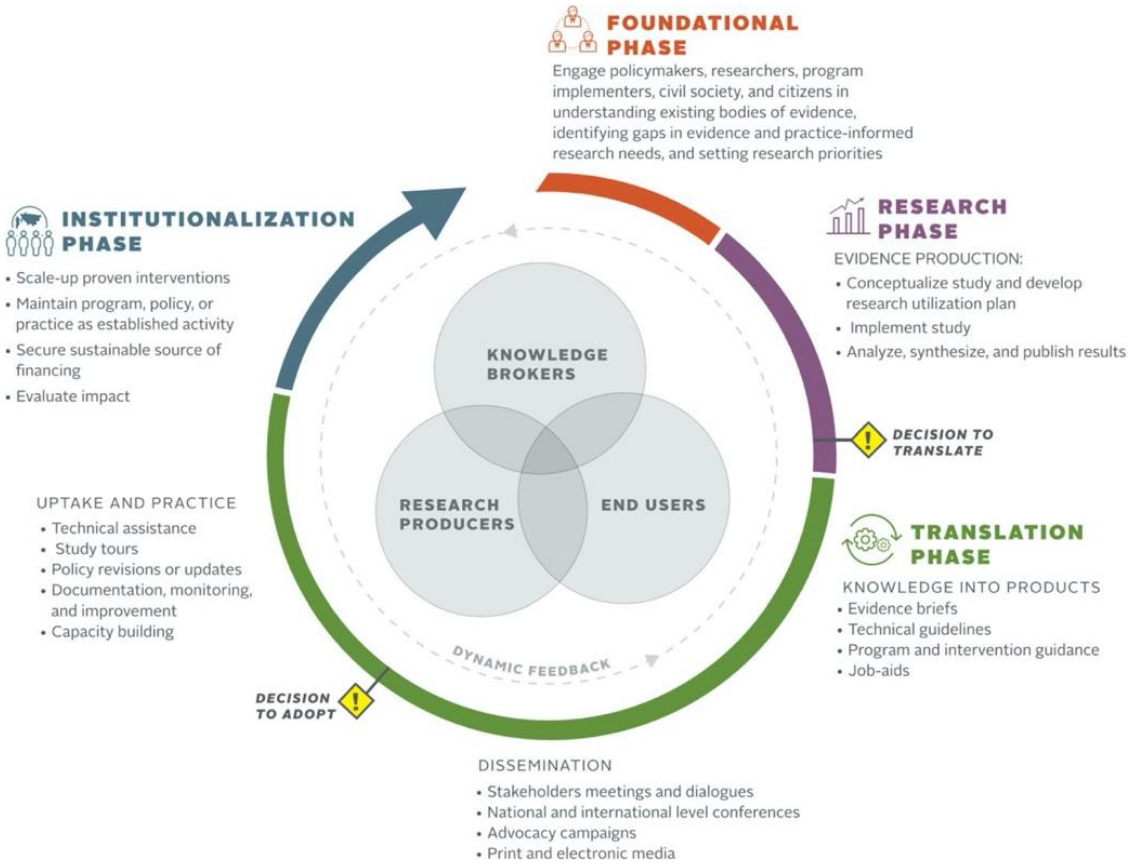


**Figure from A research utilisation framework for informing global health and development policies and programmes by Kim, et al., (2018) used under a CC-BY 4.0 licence.*

Box 7 Key Actors in Research Utilisation

Evidence Producers

Evidence producers design and execute the research.

Knowledge Brokers

A knowledge broker provides a link between researchers and research users, facilitating knowledge exchange. They foster connections across organisations and facilitate the identification, access, interpretation, and translation of research (Dobbins *et al.,* 2009). An example of a knowledge broker is FUSE, The Centre for Translational Research in Public Health (http://www.fuse.ac.uk).

End Users

End users are those who are expected to apply evidence e.g., policymakers.

The foundation phase places importance on engaging a variety of stakeholders to understand their knowledge needs and to set research priorities. This phase is instrumental to establishing ongoing relationships that will continue into the rest of the phases. The focus of stakeholder engagement aligns with the interactive model of knowledge transfer, where relationships and networks are key for research use by policymakers (Weiss, 1979).

The research phase consists of conceptualisation of the research study, conducting the research, and interpreting the findings. Although researchers play a major role in this phase, all key actors should be involved, including in the study conceptualisation as this ensures the research is relevant. This can maximise the eventual uptake of the findings. At the end of this phase, there is a critical decision point: decision to translate. This is an active decision to develop dissemination plans based on the research findings. This should be a collective decision taken by all key actors as this helps bridge the research and translational phases.

The translational phase begins to turn the research findings into actional products that can be widely disseminated and used. Evidence producers and knowledge brokers should work together on this to ensure the right strategy is taken. As described in section 1.5.4, research should be presented in short, solution-oriented summaries as this is more effective at achieving research impact. Knowledge products can be disseminated through forums, conferences, advocacy campaigns, and media engagement (Kim *et al.,* 2018). During this phase, the person translating the message is important, as well as the

target audience, as this dictates the most effective strategy for promoting research uptake.

The second critical decision point is the decision to adopt the research into policy and practice. This is made by the end users, who may have additional questions that require further investigation. This can lead to a feedback loop to the research phase before this decision to adopt can be made.

Once the end user has decided to adopt the research, it must be applied to real-life scenarios. This can be supported by evidence producers and knowledge brokers. Testing and refining of the proposed change are needed, as well as training for those implementing the change and capacity building. Documenting, monitoring, and evaluating this implementation is crucial.

The institutionalisation phase is the main outcome of the research translation process and is where the change that resulted from the research evidence is maintained. For successful implementation of health research, it needs to be sustained long-term. A number of factors influence the sustainability of health interventions and programmes such as human resources, financial resources, and ongoing political support. It also requires continuous effort to ensure the evidence is truly institutionalised.

As the FHI 360 Research Utilisation Framework considers the range of actors and activities involved at each stage of research translation process, aligning with the more complex knowledge translation theories such as the interactive model, it can be helpful for understanding how linked data research can be used by local early years' decision-makers. It also allows for some of the mechanisms for achieving research impact with linked data to be explored. Hence, this framework has underpinned the research conducted in thesis. Specifically, the ideas underpinning the foundational and research phases of the FHI 360 Research Utilisation Framework were most relevant within the scope of this thesis, meaning stakeholder engagement and collaboration is the key part of my research.

1.6 Chapter summary

This chapter described how linked routine data are being increasingly used for health research. This thesis uses the term routine data to refer to the use of data that are routinely collected and recorded electronically by public services in the UK. This includes both health and non-health data sources. There has been significant interest and

investment into linking routine data safely and securely, which has the potential to be used for research into the early years of life. This is important given the significance of the early years of life to health across the life course.

Despite the clear rationale for linking data for the purpose of improving early life health outcomes, it is unclear whether linked data research is translating into policy and practice change. This does not necessarily mean that linked data research is not informing early years policy, as this could reflect the challenges of measuring research impact. Alternatively, the challenges of using linked data, as set out in section 1.2.2, could be influencing the ability of linked data research to inform policymaking. For example, the time delays in accessing these data could make the research findings less relevant to the fast-paced policy context. If linked data are not being used to inform early years policy, this could be a missed opportunity for better informed decision-making. This is important as members of the public are only supportive of linked data research if there is transparency around how this research has translated into policy and practice (Goldacre and Morely, 2022).

The gap between research findings and policy/clinical practice is well documented and a range of interventions has been developed to increase the implementation of research beyond academia. This includes increasing support for co-produced research, where decision-makers are included at the early stages of the research cycle. It is also important that researchers invest time in understanding the policy process if they are to successfully achieve research impact (Cairney and Kwiatkowski, 2017). The *'research-policy gap'* is apparent in the literature for most scientific research, however, it is important to understand if linked data faces any specific challenges (Oliver et al., 2014). For example, there may be particular ethical and governance challenges, time delays, and data quality issues, which can potentially impact its use by policymakers. There is also limited empirical evidence around how effective the current strategies are at improving the use of evidence by decision-makers. Authors in this field have recommended further research to evaluate how different ways of engaging policymakers can achieve impact. This has informed the direction of the research conducted in this thesis.

The theories around the relationship between research, policy and practice are helpful for understanding how research (including linked data research) can inform policy decisions and the activities that might support this. In conducting the research detailed in this thesis, I assume that decision-makers are influenced by many factors when making decisions and that linked data may represent one of the sources of evidence. This makes it challenging to measure the impact of single research studies that utilise linked data. My views towards knowledge transfer align with the more complex models

of knowledge transfer, such as the interactive model and the *'enlightenment model'*, which consider the role a variety of actors in transferring ideas from research to policymakers (Weiss, 1977, 1979, 1982). I use these theories as a lens to understand how linked data can contribute to local early years decision-making and how this can be supported through research-policy engagement.

As discussed, the FHI 360 Research Utilisation Framework considers the complexity of the research-policy relationship, including the actors and activities involved in translating research into policy. Therefore, it can provide a guide for exploring how linked data research can be translated into local policy and practice. This research will focus on the early stages of the framework such as identifying and prioritising research areas for linked routine data and generating this research in a way that can support decision-making.

I also recognise that there are likely cultural differences between policymakers and academics, as proposed by the *'two-communities model'* (Caplan, 1979). These differences create an additional barrier for researchers to overcome when disseminating their research findings to policy audiences. This was apparent in the literature that explores the factors affecting research uptake (Innvaer et al., 2002; Orton et al., 2011; Oliver et al., 2014), which discuss how contact, collaboration, and relationships between policymakers and researchers is a facilitator for research use. This also aligns with Boswell and Smith (2017) theory of *'knowledge shapes policy'*, where research can be relevant to policy but that communication problems hinder its impact. Thus, practical steps can be taken to improve the flow of knowledge from research to policy. Tweed *et al.*, (2022) who examined the literature regarding the barriers and facilitators of cross-sectoral use of secondary data (which included linked routine data), found that a mutual lack of understanding between researchers and public authorities created challenges. They identified a broader theme that policymakers views on the use of data and evidence in the policy process is neglected. Tweed *et al.,* (2022) also reported there was often a mismatch between the evidence produced by academic researchers and the needs and preferences of decision-makers.

Based on the theories and literature presented, Chapter 2 of this thesis explores the evidence regarding the use of linked data by early years decision-makers in the UK and the factors that influence this. The subsequent research in this thesis explores how linked routine data can be used as a local health intelligence tool for child and maternal health, addressing the knowledge gaps identified in Chapter 2. The learning health system cycle (Figure 3), the data-driven decision-making model (Figure 4), and FHI 360 Research

Utilisation Framework (Figure 5) provide a basis for the model that underpins how I explore this. This will be discussed in more detail in Chapter 3.

Chapter 2: The use of linked data evidence in UK decision-making related to early life health: A systematic mapping review

2.1 Introduction

In Chapter 1, I discuss how health-related research utilising routine data from multiple sources is becoming increasingly widespread, including in the UK. For example, a number of birth cohort studies have sought to link routinely collected data for mothers and babies, such as BiB and the MCS. This creates opportunities for linked data to be used to explore issues around early life health and inform policymaking. Currently, it is unclear the extent to which linked data research is being used to inform early years policy and practice, how much of it can be used, and with what level of success it is currently being used.

This chapter presents a mapping review which aimed to understand how linked data research has previously been used to inform UK decisions around early life health and explores the factors affecting the use of linked data as evidence in these decisions. The purpose of this review was to identify gaps in the literature to inform subsequent research for this PhD. An additional objective was to provide a useful reference for those looking to invest in linked data research.

Section 1.5.2 presented theories of knowledge transfer, where it is acknowledged that policymakers and practitioners are likely influenced by many factors when making decisions and that research is only one of these factors. This creates challenges when measuring research impact as it is unclear what information contributed to a particular decision. In light of these challenges, the purpose of this review was not to identify all occasions where research produced using linked data have informed policy and practice decisions. The purpose was to explore where the use of linked data research by decision-makers has been reported, categorise the types of influence linked data research has achieved, and to determine the factors contributing to this successful influence.

This chapter is largely based on the review protocol which was registered on Figshare on 21st June 2022 (https://doi.org/10.6084/m9.figshare.20109593.v1).

2.2 Aim and objectives

The overall aim of this review was to provide a description of the evidence base related to the use of linked data research in early years decision-making and to identify knowledge gaps to inform further research.

In this context, the review's objectives were:

- To map the different ways in which linked data research has been used to inform UK decision-making around early life health.
- To identify strategies used to promote the use of linked data research findings and, where appropriate, evaluate the success of these strategies.
- To identify the methods used to assess the impact of linked data research.
- To explore barriers and enablers to using linked data research as evidence to support decisions around early life health.

2.3 Methods

Systematic reviews are often regarded as the gold standard method for synthesising evidence from existing primary research, either qualitatively or quantitatively, to answer clearly defined research questions (James *et al.,* 2016). However, some questions posed by practitioners and policymakers do not readily translate into topics that can be answered through combining the results in the way that a systematic review does, especially where there is limited empirical data. As a result, alternative methods of gathering and collating evidence have been proposed.

Mapping review methods were developed and refined by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) to collate, describe, and catalogue the available evidence (e.g., primary, secondary, theoretical, economic) relating to a particular topic and to identify gaps in the research literature from which to commission further research (Grant and Booth, 2009; James *et al.,* 2016). Mapping reviews follow the same rigorous, transparent, and objective processes as a systematic review to capture the available evidence relevant to a particular topic, whilst avoiding the potential pitfalls associated with traditional literature reviews. These pitfalls include lack of transparency, replicability of review methods, and comprehensiveness, as well as selection bias, where included studies are not representative of the evidence base. Table 6, adapted from James *et al.*, (2016), compares a systematic review with a mapping review.

Mapping review methods are particularly valuable when there is a wide range of evidence for a particular topic, as was anticipated in this review. For example, evidence relating to the use of linked data research could be:

- A reference to linked data research in a policy document;
- An impact case study produced by a research team that demonstrates how linked data research has been used; or
- A section of a journal article explaining how linked data research had influenced change.

This type of evidence would be challenging to synthesise using traditional systematic review methods. Mapping reviews rarely extract and synthesise study results or undertake critical appraisal due to the variety of information sources. As such, a mapping review represented an appropriate vehicle for considering a wide range of evidence on how linked data research has influenced decisions around early life health. A mapping review also presents the results in a user-friendly format, often as visual figures or a searchable database, which can be valuable for those seeking to understand how to improve the use of linked data in early years decision-making.

Table 6 Comparing a mapping review with a systematic review

Stage in evidence synthesis	Mapping review	Systematic review
Objective	Describes existing knowledge for a question or topic.	Aims to address questions that can be answered qualitatively or quantitatively.
Question formulation	Question and topic can be broad or narrow.	Usually close-framed question.
Search Strategy	Accommodates a wide range of evidence in the review.	Evidence is limited to primary qualitative or quantitative research.
Article Screening	Documents with limited data or where full text is not available can still be eligible for inclusion.	Full text is required.
Data extraction	Description of study and methods are extracted but study results are not routinely collected.	Describes the study and its methods. Extracts qualitative and quantitative results.
Critical appraisal	Critical appraisal not routinely conducted to accommodate the variety of information sources.	All included studies are critically appraised for internal and external validity.
Synthesis	No formal synthesis of study results but trends in the literature, knowledge gaps and knowledge clusters identified.	Qualitative and quantitative results are synthesised using the appropriate methodology and knowledge gaps are identified.
Report	Describes and catalogues the available evidence on a pre-specified topic, identifies gaps in the knowledge and knowledge clusters. Outlines the implications for policy, practices and/or research.	Presents narrative and qualitative or quantitative synthesis of study results addressing the review question, identifies knowledge gaps and describes implications for policy and practice.

**Table adapted from James et al., (2016).*

Similar to a mapping review, a scoping review conducts a preliminary assessment of the size and scope of the available research literature and characterises the quantity and quality of the literature (Grant and Booth, 2009). However, mapping reviews can be distinguished from scoping reviews by their output. The systematic map produced as part of a mapping review provides an explicit and transparent means of identifying specific research questions and an overall description of the evidence base in the form of a searchable database. I chose a mapping review method over a scoping review as I wanted to map out and categorise the existing literature into the different types of decisions that data linkage has informed, to inform the primary research conducted in this thesis, rather than identify the size and scope of the evidence available.

The methods used in this mapping review have been guided by Booth *et al.,* (2016), Grant and Booth (2009) and James *et al.,* (2016), following current best practice with regards to this method.

I was the lead reviewer coordinating and conducting this review. I was supported by my supervisory team (Professor Kate Pickett, Professor James Wilsdon and Sally Bridges), Information Specialists at the University of York and the University of Sheffield, and a second reviewer, Louise Padgett (LP). LP conducted 10% of the screening at both the title and abstract and full text stages to ensure accuracy during this process. Information specialists were able to advise on the search terms used and methods of searching the data sources.

Further to this, I contacted a number of experts working on linked data research projects to support children and families. These experts included senior individuals working on studies linking routine data about children and families to bespoke survey data, that were identified by my supervisors, or were identified as key authors in the field during the scoping search.

Members of my supervisory team utilised their personal contacts and introduced me to these experts by email. I then provided the experts with details of the review and asked if they could identify any relevant published documents or examples that aren't necessarily documented. I also asked for their expertise in relation to my search strategy, where they assisted with the identification of additional evidence sources and case studies not found in the systematic search. Consulting with these experts helped to ensure the mapping review was comprehensive and represented a variety of evidence on the topic. These experts also gave a clearer sense of how linked data research may inform decision-making, as the evidence relevant to this review was challenging to

identify. My reflections on consulting with these individuals are detailed in section 2.4.4, and are interpreted alongside the narrative review of the findings in section 2.5.

2.3.1 Data sources

A wide range of sources were explored as the evidence sought in this review can be found in a variety of places including academic articles published in peer reviewed journals, policy documents found on government websites, and impact case studies likely found on the websites of studies utilising linked data.

The following research databases were searched for peer reviewed literature published between January 2000 and July 2021:

- Medline (via Ovid),
- Psycinfo (via Ovid),
- Embase (via Ovid),
- Cumulated Index to Nursing and Allied Health Literature (CINHAL)(via EBSCO),
- Web of Science,
- Applied Social Sciences Index and Abstracts (ASSIA) (via Proquest),
- British Nursing Index (via Proquest), and
- Scopus.

These databases were chosen due to their relevance to health and/or data linkage research. As much of the evidence sought in this review is likely found in policy documents or on websites of studies linking data, sources of grey literature were also searched. These sources are detailed in Table 7. Grey literature is a term used to describe information produced outside of traditional publishing and not well represented in indexing databases (University of Exeter, 2022). Table 7 divides the sources into four subsections: electronic databases; websites of UK data linkage organisations who produce or facilitate linked data research; websites of policy/practice/voluntary organisations in the field of early life health who are the potential users of linked data research, and other sources. This allowed different search strategies to be applied depending on the function and purpose of the organisation.

Table 7 Alternative literature sources

Electronic databases	Websites Data linkage study/service	 Policy/practice/voluntary organisation	Other sources
Dimensions https://www.dimensions.ai	Administrative Data Research (ADR) UK https://www.adruk.org	Department of Health and Social Care (UK) www.dh.gov.uk	Google search engine (limited to the first 10 pages)
Health Management Information Consortium (HMIC) Via Ovid	Avon Longitudinal Study of Parents And Children (ALSPAC) http://www.bristol.ac.uk/alspac/	Institute of Health Equity https://www.instituteofhealthequity.org/home	Google Scholar (limited to the first three pages)
Overton https://www.overton.io	Born in Bradford (BiB) https://borninbradford.nhs.uk	ESRC https://esrc.ukri.org	Hand search reference lists of relevant documents.
	The Census and Administrative data LongitudinaL Studies Hub (CALLS-HUB) https://calls.ac.uk/output-entry/census-administrative-data-longitudinal-studies-hub/	National Institute for Health and Care Excellence (NICE) https://www.evidence.nhs.uk	Research Excellence Framework impact case studies https://impact.ref.ac.uk/casestudies/
	The Cohort and Longitudinal Studies Enhancement Resources (CLOSER) consortium. https://www.closer.ac.uk	Nuffield trust https://www.nuffieldtrust.org.uk	
	Clinical Data Linkage Service (CDLS) https://www.maudsleybrc.nihr.ac.uk/facilities/clinical-record-interactive-search-cris/cris-publications/2019/	Royal College of Paediatrics and Child Health https://www.rcpch.ac.uk	
	Clinical Practice Research Datalink (CPRD) https://www.cprd.com/services	Scottish Health Informatics Programme (SHIP) http://www.scot-ship.ac.uk/publications.html	
	Growing Up in Scotland (GUS) https://growingupinscotland.org.uk		
	Generation Scotland https://www.ed.ac.uk/generation-scotland		
	Secure Anonymised Information Linkage (SAIL) https://saildatabank.com		

Electronic databases	Websites		Other sources
	Data linkage study/service	Policy/practice/voluntary organisation	
	Understanding Society, The UK household longitudinal study https://www.understandingsociety.ac.uk		
	UK data service https://ukdataservice.ac.uk/impact/		
	Wirral Child Health and Development Study (WCHADS) https://www.liverpool.ac.uk/population-health/research/groups/first-steps/		

The list of sources in Table 7 was compiled by consulting academics in the field of data linkage and through scoping searches to ensure no relevant sites were missed. UK data linkage studies that do not address research relating to early years were not included in this list. A scoping search initially included more data linkage studies and websites, however, those that returned no relevant results or had no relevant pages to navigate were excluded (See Appendix A for additional searches included at the scoping stage).

2.3.2 Search terms

Numerous methods have been developed to assist with the formulation of search strategies. The JBI recommends the PCC mnemonic (Population, Concept and Context) for developing search strategies in scoping reviews and it is standard practice to use the Population, Intervention, Comparison, Outcomes and Context (PICOC) framework in systematic reviews (Petticrew and Roberts, 2006). In this review, relevant words associated with the PCC of the review were searched.

I considered including terms associated with the *'outcomes'* in this review, which would include terms related to decision-making with linked data research, strategies promoting the use of linked data and barriers and enablers to using linked data research. However, not including these terms allowed the search to be more sensitive to the range of possible outcomes. This is appropriate given there are a wide range of potential decisions that can be taken as a result of linked data research and these outcomes vary in the way they are reported between documents. For example, a policy document might reference linked data research in relation to a commissioning decision, or it could be a sentence at the end of a linked data research paper that explains how a decision-maker has used linked data as evidence to recommend the implementation of a service. In both cases, it

is challenging to translate these outcomes into a list of searchable terms that capture all the potentially relevant outcomes. A pilot search was conducted using only the PCC, and this produced a manageable number of results, allowing for this more sensitive search strategy.

It was pragmatic to focus on decision-making related to the early years of life, due to the importance of investment during this period for health across the life course, and this was the focus of this thesis. It was also unclear whether linked data have been used to inform decisions in this area. I considered narrowing this further to the first 1,001 days of life, however, as the evidence related to the use of research by decision-makers is challenging to identify with this degree of granularity, I wanted to include more opportunities to identify research impacts.

Expertise was sought from the Department of Health Sciences Information Specialist at the University of York, and previous authors reviewing literature in this field to develop and refine these search terms. The search terms were piloted to ensure known relevant papers appeared in the search results and the terms were amended accordingly. Table 8 presents the finalised search terms for electronic database searches through Ovid. The descriptors were translated into the appropriate format for each database interface, but the free text terms remained the same.

Table 8 Search Terms

Population	Concept	Context
"Early years".tw	"Data link*".ti	UK.tw
Infant.tw	"Link* adj3 data*".ti	"United Kingdom".tw
Infancy.tw	"Integrated data*".ti	"Great Britain" .tw
Baby.tw	"Connected data".ti	England.tw
Babies.tw	"Record link*".ti	Scotland.tw
Toddler*.tw	"information link*".ti	Wales.tw
Preschool.tw	"linked electronic health	"Northern Ireland" .tw
"First years".tw	record*".ti	Welsh.tw
"Early childhood".tw	"electronic birth cohort".ti	Scottish.tw
Child*.tw	"e-cohort".ti	English.tw
Antenatal.tw		Britain.tw
Postnatal.tw		British.tw
Newborn.tw		NOT "New South Wales".tw
Mother*.tw		
Parent*.tw		
Family.tw		
Families.tw		
Maternity.tw		
Maternal.tw		
Father.tw		
Paternal.tw		
Pregnancy.tw		
Pregnant.tw		
Perinatal. Tw		
Paediatric.tw		

*truncates the word allowing multiple variations of the term.
Tw denotes text word, meaning only titles and abstracts will be searched for each search term.
Ti denotes title, meaning only titles will be searched for those words/phrases
Adj3 allows up to three words to be in between link* and data*.

Despite some databases having the functionality to filter based on country, the diversity of the sources searched in this review meant that context terms were needed to help focus the search, in the absence of a country filter. Hence, these terms were included for consistency across sources.

The functionality of bibliographic databases allowed Booleen operators to define how each database combined the individual search terms detailed in Table 8. The search was limited to English language and documents published after 2000, where possible, to accommodate the time and resources available. All search terms were limited to titles

and abstracts to ensure the results were relevant and manageable within the scope of this review. Where possible, concept terms were limited to title searches in electronic databases, to ensure relevance of the results with respect to the review objectives. In databases that search for policy documents, the concept terms were not limited to title searches, as eligible policy documents likely include references to linked data research throughout, with the focus of the document being on early life health. In those cases, the title may not include data linkage terms. Details of the full search strategy can be found Appendix A.

As much of the literature sought in this review is reported outside of bibliographic databases and found on websites and online sources, this presented challenges for executing complex search strategies, exporting search results, and applying a consistent search strategy across sites. Stansfield *et al.,* (2016) propose a systematic approach to searching and reporting websites and online resources, which was followed in this review.

To search each online source, I made use of generic search functions, navigated headings within webpages, and scanned reference lists, where applicable (Stansfield *et al.*, 2016). All items listed under key findings on data linkage study/service websites were browsed, where publications were numbered less than 100, and advanced search features were utilised to restrict the number of documents. As there was no function to export the results of website searching into a citation management tool, I assessed the results of those searches for relevance. This involved scanning titles and abstracts, where available, of the first 100 items so that only potentially relevant items were exported. I acknowledge that there was an element of reviewer judgement applied at this stage. Alongside navigating pages that document key study findings, pages where engagement strategies for data linkage organisations and impact case studies are likely detailed were also searched, to ensure that evidence pertaining to the review objectives was identified.

The method of searching each individual source was recorded using the following headings:

- Name of resource,
- Date searched,
- Pathway followed (e.g., browsed headings, used search function within the website, browsed references),
- Key words used, and
- Number of results.

This ensured a consistent approach across online sources, enabled the flexibility to search each source differently, and the ability to search each source multiple times for different types of evidence. The results of this method can be found on the online open access repository, Figshare (https://doi.org/10.6084/m9.figshare.24005949.v1).

Complementary search approaches such as asking experts working in the field of data linkage and citation searching of relevant documents accompanied the above search strategy.

2.3.3 Eligibility criteria

An eligibility criterion related to the population, concept, context, and outcomes of the review was used to screen the identified papers for relevance. The criterion was informed by known relevant papers identified in the scoping search and is detailed in Table 9.

The search strategy identified documents that discuss linked data research related to early life health and child development within the context of the UK. By including a criterion related to outcomes at the screening stage it allowed documents that did not discuss linked data in relation to decision-making, strategies to promote the use of linked data, or the barriers and enablers to using linked data research to be filtered out. This was possible as reviewers had access to the context of the document and were able to search for evidence of how linked data might have been used. There were no restrictions on the types of evidence sources included in the review, to reflect the diverse nature of the evidence available. This is appropriate given there will be no synthesis of study results.

Table 9 Mapping review eligibility criterion

	Inclusion	Exclusion
Population	1. The decision that was informed by the linked data research must relate to the early years of life. i.e., the decision taken must have implications for children in the early years of life (from conception to age of five years) or for the parents of children in their early years, where the child is also impacted. I acknowledge that linked data research may explore how exposures in the early years of life are associated with outcomes later in life. Thus, if the decision taken as a result of the research relates to changes made for children in the early years, then it is eligible for inclusion. For example, a linked data research study may show that an intervention in early childhood is associated with positive long-term outcomes, which in turn informed a decision to commission this intervention in early childhood. This would be eligible for inclusion. 2. Where the document refers to multiple exposures and outcomes, at least one should be relevant to children under the age of five years. The mean or median age of children in the cohort study may be used to determine if the linked data research refers to children under five years. 3. If the document is referring to strategies employed to promote the use of linked data research findings or barriers and enablers to using linked data research, the research must have implications for policy related to the early years period. 4. The strategies outlined must be relevant to the use in early life decision-making or have the potential to.	1. The linked data research that does not explore exposures, outcomes or interventions related to the early years of life or does not have implications for children in the early years of life.

	Inclusion	Exclusion
Concept	1. The information used by early years decision-makers must relate to findings from linked data research. 2. In the context of this review, the linked data must join routinely collected health data with at least one other health or non-health related data source to investigate topics around early life. Examples of eligible data linkage include: ▪ Routinely collected health data linked with data on education or social care. ▪ Routinely collected health data joined with a bespoke research dataset such as survey data, observational data, or data from a randomised control trial. ▪ Maternity records linked with health visitor and/or general practice data. 3. If the document is reporting on the strategies to promote the use of research findings or the barriers and enablers to influencing policy and practice around child health and development, it must be related to the use of linked data research. 4. Eligible documents that help identify barriers and enablers to using linked data research could include qualitative research exploring the perspectives of stakeholders towards linked data. 5. It would be advantageous if the document described how this influence was achieved, however studies simply stating the influence of their findings outside of academia are considered eligible.	1. If the findings influencing the decision are not related to linked data research, such as findings relating solely to cohort survey data.
Context	1. The data linkage study population and subsequent decision-making must be within the context of the UK. 2. A document is eligible if it makes cross-country comparisons, providing it details the impact of the linked data research on UK decisions related to early life health.	1. The data linkage study population and findings relate to a population outside of the United Kingdom.

	Inclusion	Exclusion
		2. The decision made as a result of the data linkage findings effects a population outside of the United Kingdom.
Outcome	The document must detail one of the following outcomes: 1. The document must demonstrate how the findings of a data linkage study have influenced decisions related to early life. This could include changes to policy, clinical guidelines, clinical practice, voluntary sector practices or wider behavioural changes at the local or national level. 2. Documents that indirectly mention how linked data study findings were used to influence policy and practice are eligible despite this not being the focus of the paper. 3. Describe the strategies used to encourage the use of linked data research in early life decision-making. 4. Describe the barriers or enablers to influencing policy and practice related to early life for linked data research (researcher perspective) 5. Describe the barriers or enablers to using linked data research as evidence in decision-making (decision-maker perspective).	1. The empirical study does not reference how the findings influenced early life decision-making.
Evidence source	1. There are no restrictions on the types of evidence sources included. 2. Documents must be published after the year 2000. 3. Documents must be published in English language.	1. Studies published before the year 2000. 2. Documents not written in English language.

The search was narrowed to the UK as my research focuses on the use of routinely collected data from public services in the UK for the purposes of informing local decision-making around child and maternal health. Given the variation in data governance laws, as well as the nature, provision and commissioning of health and care services between countries, it would be difficult to use information from different countries to inform this research.

When exploring the strategies promoting the use of linked data research, those that had the potential to inform early years policy were judged relevant, as these documents rarely reported the age of the population. Thus, by excluding all documents where age is not referenced, it would reduce the knowledge gained in this review. Hence, if the data linkage population included children in the early years of life and the strategies could be applicable to evidence related to this population, then the document was eligible for inclusion.

2.3.4 Evidence Selection

All documents identified by the search strategy were exported to the reference management software *'Endnote X9'* and all duplicates were removed. The results from the web searches were first exported into Paperpile, for efficient referencing and then exported into *'Endnote X9'*. All documents were then imported into the software package Rayaan.ai to support the screening process.

To ensure accuracy during this process, two reviewers (myself and LP) independently screened the titles and abstracts of 10% of the documents against the eligibility criteria. To randomly assign 10% of the full texts for this validity check, I organised them in *'Endnote X9'* by title in alphabetical order and assigned every 10th article for screening. It is good practice to have 100% of the documents screened by a second reviewer, however, due to capacity constraints of this project, 10% was deemed appropriate. Inter-rater reliability was then assessed using Cohens Kappa score to determine whether strong agreement score of 0.8 or greater was achieved (McHugh, 2012). A Kappa Score of 0.94 was attained and disagreements were resolved through discussion. I screened the remaining titles and abstracts. This process was reapplied at the full text screening stage, where a Kappa score of 1.0 was achieved.

2.3.5 Data extraction

A mapping review catalogues the evidence gathered usually in the form of a database, which can provide detailed *'meta-data'* about each of the included documents and their source.

A data extraction form was developed based on the information needed to address the review objectives and converted into Google Form. Relevant information included:

- Data relating to the evidence source (i.e., author, year and whether the source is peer reviewed or grey literature),
- Details of the data linkage (i.e., population whose data are linked, purpose of the linkage, what information is linked and whether the linkage is local, regional, or national),
- Details of the child health/development outcome that the data linkage findings relate to,
- Details of how the findings were used (i.e., who used the findings, what decision did they inform, how was this achieved and how was it measured),
- Data on any strategies employed to encourage the use of the data linkage findings (i.e., methods of stakeholder engagement and whether these were successful),
- Details of any barriers and enablers to using linked data research that were discussed.

The data extraction form was then used to populate the systematic map of pre-defined categories. Categories included full reference, data sources linked, child health outcomes and type of research impact/influence. The systematic map is available on Figshare (https://doi.org/10.6084/m9.figshare.24006906.v1). This database can be used by linked data researchers and those investing in linked data research to understand the available evidence pertaining to the use of linked data research related to the early years beyond academia.

The data extraction form was piloted using the first three included documents and refined accordingly before being applied to all eligible documents. I extracted the data from all the included studies and a second reviewer (LP) extracted the data from one of the included studies, to ensure accurate data extraction. Any discrepancies in the data extraction were resolved through discussion.

2.3.6 Quality Assessment

As outlined in Grant and Booth (2009), mapping reviews do not usually include a quality assessment process. The type of evidence this review considers is the type for which

critical appraisal is not appropriate. As such, the purpose of the documents eligible for inclusion may not have been to evaluate the impact of the linked data research findings but to simply present the results of the linked data research with a small section on how the findings were used.

2.4 Results

2.4.1 Total Studies

The search strategy identified 2,400 potentially eligible records once duplicates were removed. A total of eight documents were included in this mapping review following full text review. Figure 6 presents the results of the systematic search in a flow diagram.

Figure 6 Flow chart of search strategy adapted from PRISMA diagram*

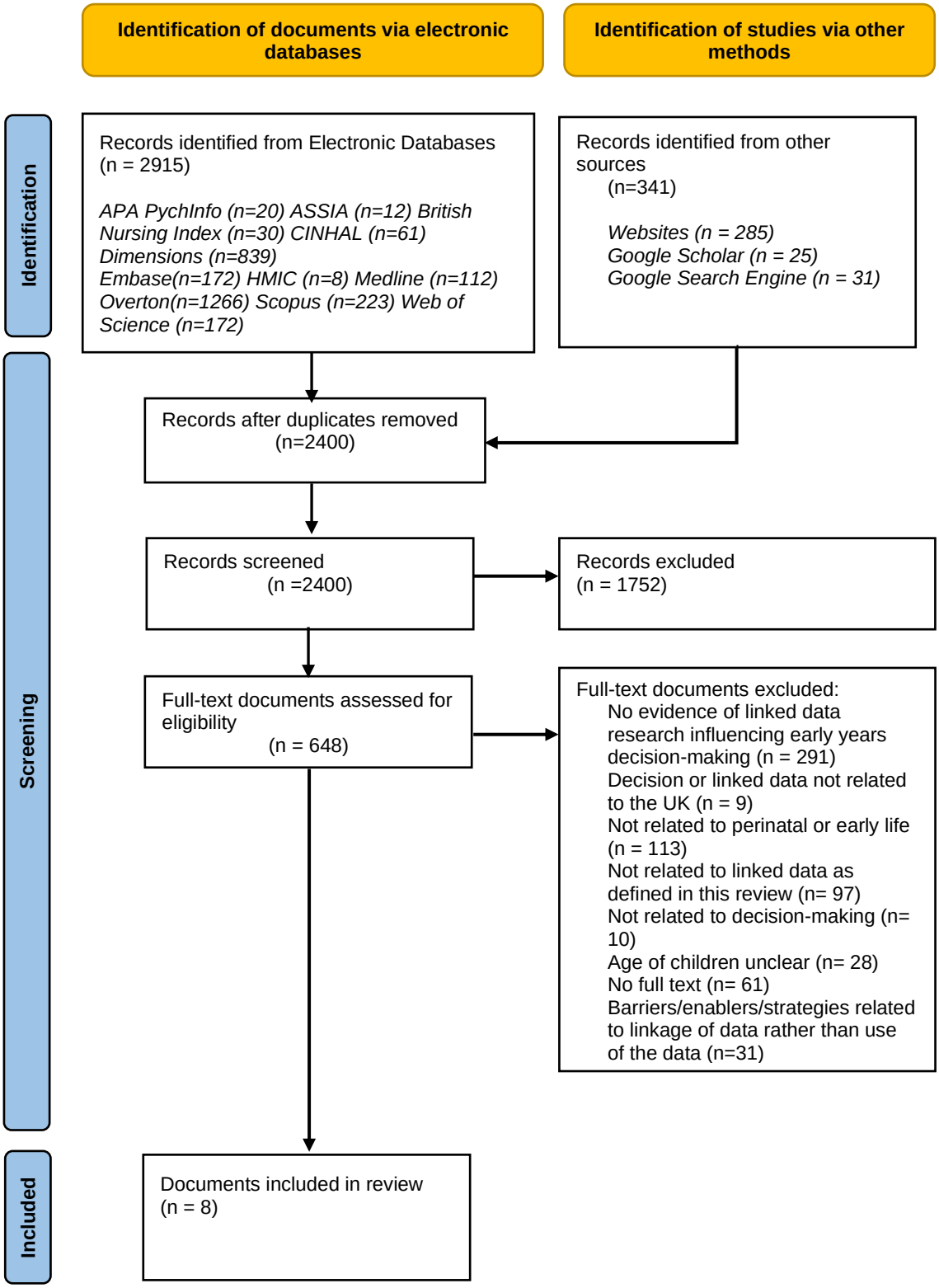


**Figure adapted from 'The PRIMSA 2020 statement: an updated guideline for reporting systematic reviews' by Page, et al., (2020) used under a CC-BY 4.0 licence.*

The initial 2,400 documents identified by the search strategy included policy documents related to the early years of life and empirical studies utilising linked data. As discussed previously, there is an abundance of linked data research in the UK, therefore, it was important for this review to capture this research, so that reviewers could determine whether it had influenced decision-making.

Many of the studies excluded under the reason *"No evidence of linked data research influencing early years decision-making"* describe linked data research but this research cannot clearly be linked with UK decision-making around the early years of life.

Some of the search results were weblinks that led to multiple documents, where each document was screened for relevance.

2.4.2 Distribution of studies

All documents were published since 2010. This accords with the increased use of linked data for research over the last decade.

Table 10 demonstrates the distribution of studies based on the geographical area covered by the data linkage. A variety of nations were covered in the documents identified, where no documents were specifically related to Northern Ireland.

Table 10 Distribution of documents based on area covered by linkage, where reported

Study location	Number of included documents
Scotland	2
England	2
England and Wales	1
United Kingdom	2

Most of the identified documents were discovered during the grey literature searches, with reports being the most common document type, as shown in Table 11. Three of the included studies were published in the peer-reviewed literature. Eligible documents were more likely to be found in the grey literature, as research studies often don't know the impact of their research until after they have published the results in the peer review literature. Hence any impacts are likely reported on organisational websites or in reports.

Table 11 Distribution of documents based on document type

Document type	Number of included documents
Peer reviewed article	3
Report	4
Policy document	1

Table 12 represents the distribution of documents with respect to the data sources included in the linkage. One article describes multiple studies that utilise data linkage, where each link to different datasets. This was categorised under *'routinely collected health data sources and other sources'*. Most of the included documents utilised routinely collected data alongside other non-routinely collected information such as cohort observational data, survey data or non-health data. Of those that linked routinely collected health data sources and other sources, four linked to education data. Few studies (n=2) linked routine data exclusively from health data sources.

Table 12 Distribution of documents based on data sources linked

Datasets linked	Number of included documents
Routinely collected health data sources and other sources	6
Routinely collected health data sources	2

Table 13 presents the distribution of documents based on the outcome they address that is relevant to this review. Strategies to promote the use of linked data research was the most frequently reported outcome (n=6). There were a relatively small number of documents that reported the use of linked data research by UK decision-makers (n=2). Two documents addressed both strategies to promote the use of linked data research findings and barriers/facilitators to using linked data evidence.

Table 13 Distribution of documents based on the outcome relevant to this review

Outcome relevant to this review	Number of included documents
Strategies to promote the use of linked data research findings	6
Barriers and facilitators to using linked data evidence	2
Linked data influenced decision-making	2

2.4.3 Summary of the included documents

A summary of the included documents is detailed in Table 14. A list of excluded documents has been maintained and is available on request.

Table 14 Summary of included studies

Reference ID and short identifier	Reference	Study location	Data sources linked	Early life outcome	Outcome relevant to this review
1. ADR UK (2021b)	ADR UK. (2021b) *Our public engagement activities* [Online]. Available: https://www.adruk.org/our-mission/our-public-engagement-activities/ [Accessed 3rd February 2022].	UK	ADR UK host the ECHILD database which links administrative health and education data for children.	Child health and education	Strategies to promote the use of linked data research findings
2. Davis-Kean *et al.,* (2017)	Davis-Kean, P., Chambers, R., Davidson, L., Kleinert, C., Ren, Q. & Tang, S. (2017). *Longitudinal Studies Strategic Review 2017 Report to the Economic and Social Research Council*.	Includes a range of data linkage studies covering areas across the UK	N/A	N/A	Strategies to promote the use of linked data research findings, Barriers and facilitators to using linked data evidence
3. Hopf *et al.,* (2014)	Hopf, Y. M., Bond, C., Francis, J., Haughney, J. & Helms, P. J. (2014). 'The more you link, the more you risk ...' - A focus group study exploring views about data linkage for pharmacovigilance. *British Journal of Clinical Pharmacology,* 78, 1143-1150.	Scotland	Linking of routinely collected national paediatric data. Combining datasets from primary and secondary care	Paediatric Adverse Drug Reactions	Strategies to promote the use of linked data research findings
4. Macfarlane *et al.,* (2019)	Macfarlane, A., Dattani, N., Gibson, R., Harper, G., Martin, P., Scanlon, M., Newburn, M. & Cortina-Borja, M. (2019). Births and their outcomes by time, day and year: a retrospective birth cohort data linkage study. *Health Services and Delivery Research*, 7, 1-268.	England, Wales	ONS birth records, death registration records, Hospital Episode Statistics in England and Patient Episode Database for Wales linked with National Community Child Health Database. The Centre for Maternal and Child Enquiries data for England and Wales	Mortality of babies and mothers, and morbidity recorded at birth and any subsequent hospital admission. They examined patterns in births by time, day and year and variations between	Strategies to promote the use of linked data research findings, Barriers and facilitators to using linked data evidence

Reference ID and short identifier	Reference	Study location	Data sources linked	Early life outcome	Outcome relevant to this review
			from 2005 to 2009 were linked to stillbirth registration records	maternity services in relation to medical and midwifery staffing, intervention and size of unit. They examined how the outcome of pregnancy.	
5. Powell *et al.* (2021)	Powell, T., Gheera, M., Foster, D., Long, R. & Kennedy, S. (2021). *Early intervention: policy and provision*, UK Parliament Research Briefings.	England	Specifics of the data linkage not specified, however the source it references links maternally reported baseline and follow-up data with, Hospital Episode Statistics data, social care, educational data, and abortions data.	Child development at reception age.	Linked data influenced decision-making
6. Robling *et al.* (2021)	Robling, M., Lugg-Widger, F., Cannings-John, R., Sanders, J., Angel, L., Channon, S., Fitzsimmons, D., Hood, K., Kenkre, J., Moody, G., Owen-Jones, E., Pockett, R., Segrott, J. & Slater, T. (2021). The Family Nurse Partnership to reduce maltreatment and improve child health and development in young children: the BB:2 6 routine data-linkage follow-up to earlier RCT. *Public Health Research,* 9, 1-160.	England	Maternally reported baseline and follow-up data from a randomised control trial was linked with, Hospital Episode Statistics (through NHS Digital), social care and educational data (from the National Pupil Database) and abortions data (from the Department of Health and Social Care) for both the mother and child.	Primary outcome: child-in-need status, secondary outcomes: (1) referral to social services, child protection registration, child-in-need categorisation, looked-after status, recorded injuries and ingestions at any time during follow-up; (2) early childcare and educational attendance, school	Linked data influenced decision-making

Reference ID and short identifier	Reference	Study location	Data sources linked	Early life outcome	Outcome relevant to this review
				readiness (Early Years Foundation Stage Profile score) and (3) health-care costs.	
7. Growing Up in Scotland (2010)	Growing Up in Scotland. (2010). *Using the findings from the Growing Up in Scotland study - a guide for Local Authorities*, Centre for Research on Families and Relationships, The University of Edinburgh.	Scotland	Data linkage of cohort questionnaire data to administrative data	Early life topics covered include characteristics and circumstances of children and their families; pregnancy and birth; parental support; parenting styles and responsibilities; childcare and work-life balance; child health and development; parental health; food and eating; and experiences of pre-school education.	Strategies to promote the use of linked data research findings
8. Waind, and Mc Grath-Lone (2021)	Waind, E. and Mc Grath-Lone, L. (2021). *The Education & Child Health Insights from Linked Data (ECHILD) Database Key messages from stakeholders*, ADR UK.	UK	The ECHILD Database links health, education, and social care data for all children in England. It currently links two the National Pupil Database and Hospital Episode Statistics, and there are plans to link primary care data, Community Services Data Set, Longitudinal Education Outcomes, Data from local authorities.	Health and education	Strategies to promote the use of linked data research findings

2.4.4 Expert consultation

This section highlights the key messages from consulting with experts in the field of data linkage, which complement the findings from the literature searches. No exact quotations are included to protect the anonymity of those who were involved.

Before beginning the review, I consulted with information specialists working on a mapping review exploring the use of linked routine data in local commissioning settings. They were able to advise on the mapping review method, searching the grey literature, and they identified case studies where information had been shared between services that could improve commissioning of services to enhance wellbeing of communities. Unfortunately, the cases where routine data were linked to explore the early years of life outcomes provided no evidence of impact on UK decision making.

As a result of the search strategy a webpage was identified entitled *'Linking datasets for the benefit of children and young people'* (Nuffield Trust, 2022). It related to an event held by the Nuffield Trust in November 2019, which brought together key stakeholders from health and care, local government, and research organisations to learn how data linkage can be used to benefit children and young people. It also explored new insights derived from linked data and what changes had been made locally as a result. There were no details on the webpage relating to the discussions that took place at the event. Thus, to find out more information, I contacted the organisers of the event using the details provided on the page. I arranged an online meeting with a member of the team who attended the event, to discuss the event in relation to the objectives of this review. A reflection from the event was that successful data linkage projects are driven by highly motivated individuals pushing the agenda forward. It was suggested that linked data may be informing practice locally and used strategically, but details of how it has been used are unlikely to be published, unless partnered with an academic institution. In contrast, university-based researchers are likely to publish their findings in journals, as a higher priority than informing policy and practice. This could explain why few eligible papers were identified in this review, as information about how linked data have been used is not consistently reported in a discoverable way.

Another relevant discussion from this event related to the challenges of using linked data locally. For local authorities to use linked data research, it needs to be made available in an accessible format and the lack of analytical capability within local authorities is a barrier to using linked data as evidence. In addition, information systems within local authorities prevent them from easily accessing software which allows analysis of the

data, thus creating a barrier to using linked data to inform decision-making. A final reflection related to understanding the importance of linked data research in decision-making and incorporating this more formally into decision-maker roles.

Experts at data linkage studies were unable to identify well documented examples of where linked routine data research had informed policy and practice. Many of the other research papers suggested by informants which have achieved impact, are beyond the scope of this review as they were focused on educational outcomes for children over the age of five years. This lack of evidence was explained by the limited incentives to trace the impact once the research funding stream has ended. There is also no standardised way of reporting the impact of linked data research, making studies less likely to report on this in a discoverable way.

After consulting with multiple experts working on different data linkage studies and those working in this field, I reached a saturation of responses, where no additional case studies were being identified, and carried out no further consultation.

2.4.5 Narrative review

This section discusses the identified studies with respect to the review objectives.

A range of data sources are linked between each of the included documents, demonstrating the scope of linked data research. A summary is provided here:

- Two of the documents refer to ADR UK which hosts the ECHILD database, linking health and education data for children (ADR UK, 2021e and Waind, and Mc Grath-Lone, 2021);
- Two documents refer to the Family Nurse Partnership evaluation which links maternally reported baseline data and follow-up data from a randomised control trial with Hospital Episode Statistics data, social care, educational data and abortions data (Powell *et al.,* 2021; Robling *et al.,* 2021);
- One study combined datasets from primary and secondary care for pharmacovigilance (Hopf et al. 2014);
- One study links birth records, death registration records, Hospital Episode Statistics in England, Patient Episode Database for Wales and the National Community Child Health Database (NCCHD) (Macfarlane et al. 2019);
- One document links administrative data to cohort questionnaire data in Scotland (Growing Up in Scotland, 2010).

The child health outcomes reported in the included documents are wide ranging. They include developmental outcomes at preschool age, mortality and morbidity, patterns in birth timing, and paediatric adverse drug reactions.

2.4.5.1 Use of linked data research to inform UK decision-making around the early years of life

Two documents identified in this review reported the use of linked data research findings by UK decision-makers, at the national and local level (Powell *et al.,* 2021; Robling *et al.,* 2021), both of which were published in 2021. These two documents refer to an evaluation of the Family Nurse Partnership (FNP), which utilises linked routine data to follow-up an earlier randomised control trial. The evaluation was a stand-alone data linkage study and not part of a data linkage cohort study. The FNP provides support to mothers under the age of 24 from early pregnancy until their child is aged between one and two years. The peer-reviewed paper by Robling *et al.,* (2021) is the original document describing the methods and results of the data linkage study. The paper demonstrates that the FNP improves child development at age five years (Family Nurse Partnership, 2022).

It was included in this review, as in the section detailing the implications of the results on practice, it stated that the *FNP "remains locally commissioned and delivered in England. Local needs and priorities may determine the weight attached to these different sets of outcomes"* (pg.101 Robling *et al.,* 2021). As a result of the linkage study, some local decision-makers have continued to commission the service, which is provided to families during the early years.

Prior to the study by Robling *et al.,* (2021), some local areas had begun to decommission the FNP e.g., Bradford decommissioned the service in 2018 (Family Nurse Partnership National Unit & Dartington Service Design Lab, 2020). The findings from the original randomised control trial showed no evidence of the intervention improving outcomes in UK. However, the researchers involved in the Robling *et al.,* (2021) study were aware of studies in the United States showing the programmes to work in the longer term. Hence, data linkage enabled longer term follow-up, allowing additional benefits in the UK to be examined. It is unclear from the evidence presented whether local areas recommissioned the FNP or if those already providing the service continued to provide the FNP, as a result of this research.

A policy briefing by Powell *et al*., (2021) titled *'Early intervention: policy and provision',* references the data linkage study by Robling *et al.,* (2021). In the policy document, they

refer to the Health and Social Care Committee's report on the First 1,000 days of life where they recommend that the Government works with local areas and the voluntary sector to develop a programme into which children and families who need targeted support can be referred, drawing on the experience of the FNP in Scotland, Northern Ireland, and in some parts of England.

"The Committee also agreed with the Science and Technology Select Committee that commissioners should continue to appraise the evidence base for the FNP, as well as for other targeted interventions, and consider investment or disinvestment accordingly" (pg.18 Powell *et al.*, 2021).

The document describes the Government's response to this, which is outlined in their work with the Early Intervention Foundation as a *'What Works Centre'*, ensuring that investment in services is evidence-based and has a strong impact on child outcomes. The Government outlined the following in relation to the FNP:

"The FNP programme uses an approach to share learning and evidence that once tested has the potential to benefit a wider cohort of families. In April 2020, the FNP National Unit function will transfer to in-house within Public Health England to enable sustainability, significantly better taxpayer value, and dissemination of skills and knowledge across a range of high priority early years interventions. This will enable [Public Health England] to deliver the FNP National Unit functions to fulfil the FNP licence requirements for England, as well as supporting cross government priorities on the first 1000 days in order to benefit a wider cohort of children" (pg.18, Powell *et al.*, 2021).

In relation to the objectives of this review, the document by Powell *et al.*, (2021), shows the linked data research by Robling *et al.,* (2021) being used by the Health and Social Care Committee and the UK government to make decisions about the FNP. By only reading the report by Powell *et al.,* (2021), details of the data linkage are unclear. Therefore, to establish eligibility for this review, reviewers located the original document, Robling *et al.,* (2021), to ensure that it referred to the correct PCC. This process was carried out for all policy documents that referenced linked data research.

The evidence relating to the use of linked data research in decision-making was challenging to identify in both documents. In Powell *et al.*, (2021), it was a short paragraph referencing the evaluation study by Robling *et al.,* (2021). In Robling *et al.,* (2021), it was one line on page 101 that demonstrated how it had informed commissioning decisions. This demonstrates the challenges of understanding how

linked data research are being used, as the evidence is not presented in an easily discoverable way.

No other data linkage studies identified in this review reported evidence of their findings informing UK decision-making around the early years of life, nor did any policy document refer to data linkage findings within the field of early life.

No documents were identified that assessed the impact of linked data research and there was no standardised way of reporting impact for linked data research. This reveals a gap in the knowledge for whether linked data research studies evaluate the impact of their findings.

2.4.5.2 Strategies used to promote the use of linked data research findings

Six documents identified in this review suggested strategies for promoting the use of linked data research by decision-makers beyond academia (ADR UK, 2021e; Davis-Kean et al. 2017; Hopf et al. 2014; Macfarlane et al. 2019; Growing Up in Scotland, 2010; Waind, and Mc Grath-Lone, 2021). Two documents were published in the peer-reviewed literature (Hopf et al., 2014 and Macfarlane et al., 2019) whilst others were reports or webpages located by the grey literature searches. Webpages and reports are likely more common as they are easily discoverable to other data linkage studies wishing to access this information.

Across the body of included literature, different strategies for promoting the use of linked data were identified. These strategies fell broadly into three categories: engagement activities; dissemination and communication of results; and providing advice for research users.

Three of the six documents (ADR UK, 2021e; Macfarlane *et al.,* 2019; Waind, and Mc Grath-Lone, 2021) describe methods of engagement for promoting the use of linked data research findings. Three documents (Davis-Kean *et al.,* 2017; Macfarlane *et al.,* 2019; Hopf *et al.*, 2014) describe the dissemination of results to encourage use by decision-makers, where Macfarlane *et al.,* (2019) discusses how their engagement activities supported the dissemination of their research. Finally, one document (Growing Up in Scotland, 2010) provided a guide for local authorities using findings from GUS.

The webpage by ADR UK (2021b) titled *'Our public engagement activities',* does not directly reference the term linked data, however, it is widely known that ADR UK joins administrative data from across the public sector and makes it available for research in a safe and secure way to enable better informed policy and decisions. They also host the ECHILD database which links health and education data from birth to age 25 years. The strategies outlined in this document have the potential to promote the use of linked data for early years decision-making. Hence, this page was included as the reviewers agreed and were satisfied that the strategies represented have the potential to improve the use of linked data in early years decision-making. The webpage describes how ADR UK engages with charities and voluntary organisations to understand their specific research needs and empowers these groups by offering the opportunity to influence the direction and outcomes of the research. They also discuss their country specific engagement strategies which are presented in Table 15. The theme running through each of the strategies is engagement of local organisations to raise awareness of the linked data and its potential for answering locally relevant questions.

Table 15 ADR's Engagement Strategies

ADR Organisation	Engagement Strategy
ADR Northern Ireland	They hold a Data Workshop Series around themes of interest to both researchers and local organisations. These focus on raising awareness among third sector groups about the power and potential of data in their own work and how complex questions can be answered using data, as well as embedding positive working relationships with the third sector. By bringing key stakeholders onto steering committees for each of its projects, they maximise engagement with people and organisations with differing expertise and knowledge of the issues researchers are exploring.
ADR Scotland	Conduct project-specific conversations with third sector organisations able to speak on behalf of the publics and communities relevant to each of their projects.
ADR Wales	Holds stakeholder workshops with devolved and local government and third sector organisations to get feedback on work already done and gain input on future work.
ADR England	Oversees community representative panels made up of third sector representatives, practitioners and others working directly with or on behalf of particular groups. For example, the ADR England Children and Young People Representative Panel is made up of people working directly with or on behalf of children. These panels help shape ADR to deliver the greatest possible benefits for the public.

A report by Waind and McGrath-Lone (2021) focused specifically on the ECHILD database hosted by ADR UK and presented the key messages from stakeholders. Similar to the ADR UK webpage (ADR UK, 2021e), the report stresses the importance of engaging the intended users of the research, particularly in developing the research

questions. This is to ensure that the research findings are useful to them, and they can understand and apply them to the immediate benefit of children.

In contrast, a paper by Macfarlane *et al.,* (2019) describes their use of public engagement to disseminate the findings to promote the use beyond academia. They describe how key individuals from National Childbirth Trust (NCT) assisted with dissemination of findings through social media using a short accessible hashtag, and they have plans to disseminate to parents, parent representatives and NCT practitioners. They have also spoken at conferences and published in the peer reviewed literature to promote the use of their findings. However, they explain that due to the delays experienced during the project, their dissemination plans have been limited.

The difference between the strategies of ADR UK and Macfarlane *et al.,* (2019) is that Macfarlane *et al.* (2019) are focused on engagement at the end of the research study whereas ADR UK is more concerned with ensuring the research is relevant to local stakeholders from the beginning. This is likely because ADR UK has the potential to be used for a range of research studies related to the early years of life whereas the Macfarlane *et al.* (2019) study is an *ad hoc* linkage focused on a particular area of child and maternal health, which they already know to be relevant to policy.

Similar to Macfarlane *et al.,* (2019), Hopf *et al*., (2014) focuses their strategy on dissemination. Hopf *et al*., (2014) explored the views of frontline healthcare professionals on linking routinely collected national paediatric data for the purpose of identifying earlier signals of adverse drug reactions in Scotland. Participants in the study suggested modes of dissemination that included reports and publishing in the peer-reviewed literature. They considered databases like NHS e-library (now called The Knowledge Network; http://www.knowledge.scot.nhs.uk/home.aspx) as well as websites similar to NHS Clinical Evidence (http://www.cks.nhs.uk/home). The need for concise, *'short and snappy'* information was emphasised as health care professionals are already inundated with information, which then gets ignored. This is consistent with other scientific research (Oliver *et al.,* 2014). A paediatric pharmacist highlighted that this information should also be sent to someone *"who can do something about it"*, stressing that this linked data research should be targeted at those who make the decisions. As the study by Hopf *et al*., (2014) engaged local decision-makers in this research, they were also ensuring that the linked data were being used for purposes relevant to their needs, similar to the aim of ADR UK's strategies. However, Hopf *et al*., (2014) describe no further plans to engage these stakeholders beyond this initial research. The suggestions put forward by the stakeholders participating in Hopf *et al*., (2014) are useful as they directly describe the perspectives of those who have the potential to use the linked data to inform early years

decision-making. These suggestions could also be categorised as enablers for decision-makers to make use of linked data research.

Similarly, Davis-Kean *et al.,* (2017) focus on presenting and communicating the data and research to decision-makers to ensure they can make use of it. Davis-Kean *et al.,* (2017) present the outcomes from consultation with key stakeholders and experts to explore the advantages and challenges to using administrative data linkage and how training, access, and promotion of longitudinal investments can be continued and enhanced. This was part of a review of longitudinal studies funded by the Economic and Social Research Council (ESRC). They describe how one of their funded projects, CLOSER, has produced resources to promote the value of administrative data for research. The report describes services that assist with the discovery of longitudinal research by decision-makers, including linked data research, such as CLOSER Discovery and UK Data Service (UKDS) discovery site. However, it is unclear whether policymakers access these platforms. From the statistics on data downloads from UKDS, government users represent just 2% of the UKDS downloads of longitudinal study data. The report recommends constructing an administrative data spine to be used as a basis for data linkage. A data administrative spine is described as a register that contains a record of every individual in the UK population, including their contact details, and key health and social information such as health status, age, and socio-economic status. This can be used as a research tool to understand if existing studies are representative of the wider population, allowing decision-makers to understand how to interpret the findings when designing policies (UKRI, 2022b).

Finally, Davis-Kean *et al.,* (2017) recommend that the ESRC, which funds and co-funds multiple studies that utilise linked data research, support longitudinal investments to develop innovative technology, tools, methods, and measures to track and compile metrics on data use and sharing. This could enable the impact of this research to be demonstrated. This could also aid the discovery of the information sought in this mapping review. They recommend that ESRC, in collaboration with UKDS, funds the development of a centralised analysis platform, aimed at policy users. The aim is to facilitate analysis of longitudinal data, including linked data. This platform would provide descriptive statistics and share data in a way that is accessible to users who are interested in longitudinal data, but with diverse interests and varying levels of statistical knowledge and methodological training. This has yet to be implemented, therefore, I was unable to determine whether this would encourage the use of linked data research in decision-making.

The strategies reported by Davis-Kean *et al.,* (2017) likely differ from the others as they are promoting the use of multiple studies utilising linked data, therefore, a platform improving the discoverability of data and evidence is likely to have the most impact. This is because decision-makers can seek out the evidence most relevant to their needs, whereas *ad hoc* data linkage studies such as Macfarlane *et al.,* (2019) focus on getting the evidence from one study to the relevant decision-makers, which calls for targeted dissemination.

Finally, a report by Growing Up in Scotland (2010) provided a guide for local authorities using findings from GUS. Information on GUS is presented in Table 1. The fact that the report exists is evidence of a strategy to promote the use of linked data research as it explains how their data can be used to improve local early years policy and services.

There does not appear to be a best strategy for engaging stakeholders or encouraging the use of linked data by decision-makers. There were variations in the identified strategies such as ADR UK (2021b) focused on engaging stakeholders at the start, whereas Davis-Kean *et al.,* (2017) appeared to focus on how the research could be accessed once it has been produced. None of the documents identified described whether the strategies had successfully promoted the use of linked data by decision-makers, thus revealing a gap in knowledge as to how to successfully promote the use of linked data research. I was, therefore, unable to evaluate the impact of the strategies on the use of linked data in decision-making.

2.4.5.3 Barriers and enablers to using linked data research in UK decision-making around the early years of life

Two of the papers identified in this review (Davis-Kean *et al.,* 2017; Macfarlane *et al.,* 2019) discussed barriers to using linked data research in UK decision-making as well as the strategies used to promote the use of linked data research. Some of the barriers or challenges presented in these documents are presented as barriers to linking administrative data, which can also influence the use of linked data research. For example, Davis-Kean *et al.,* (2017) suggest that the challenges associated with linking longitudinal survey data to administrative data include: user training, resource, access, consent, legislation, and data management, where these factors also affect the ability to use linked data for research and decision-making. Davis-Kean *et al.,* (2017) also stipulated that the needs of policymakers and non-academics differ from the needs of the scientific community in three central aspects: first, they need results more quickly to be able to react in time and reach the affected population with policy outcomes; second, they have a large interest in observing changes over time to identify spheres of political

action; and third, they often need data from intervention studies to evaluate measures in trial studies before implementing them. The first aspect is a barrier that linked data needs to overcome if decision-makers are to make use of it as evidence in decision-making. This is in accordance with the discussions from the Nuffield Trust event presented in section 2.4.4. Therefore, it is recommended that the research be made available in an accessible and timely manner. The second and third aspects present opportunities for linked data research to inform policy and practice.

Macfarlane *et al.,* (2019), suggest that the wider implications of the unsuccessful attempt to establish the care.data system was a barrier to achieving impact. The care.data programme is discussed in section 1.2.4. The legacy the care.data programme left for data linkage could explain why new data linkages are struggling to influence decision-making. Macfarlane *et al.,* (2019) also describe issues with quality around the Hospital Episode Statistics submitted by some maternity units, which can also be interpreted as a barrier to using linked data research in decision-making, as the research may seem unreliable. Macfarlane *et al.,* (2019) suggested that Public Health Analysts being moved out of the health service to local authorities was a barrier to accessing the data and for decision-making to be informed by their research. The delays they faced as part of the project with accessing the linked data limited the impact their findings were able to achieve. This links back to the point made by Davis-Kean *et al.,* (2017) regarding the need for timely research.

Overall, the key barriers identified from these two documents are that policymakers need to be able to trust the data linkage is ethical and of good quality to inform decision-making and that the evidence needs to be timely, accessible, and policy relevant. The barriers identified in this review were implicit and more research is needed to further understand the factors that enable and prevent decision-makers using linked data research.

2.5 Discussion

2.5.1 Key findings

The evidence related to how linked data have been used by decision-makers to inform early years policy and practice is not easy to discover. This reveals a gap in the knowledge regarding how linked data research is used by these decision-makers. The only evidence identified in this review related to an evaluation of the FNP, which used linked routine data to follow-up a randomised control trial, where local and national decision-makers used this to inform commissioning decisions and discussions. Despite the data linkage study being independent and funded by the NIHR, the FNP is

government funded and policy relevant (as demonstrated in their first 1000 days of life publication). This could explain why local and national decision-makers have made decisions using this linked data evidence (House of Commons, 2019). This may be because the government were aware of the study from the start, making it more likely that they would be invested in the results and recommendations. This aligns with one of the strategies identified to promote the use of linked data research, which involved engaging with stakeholders at the start and throughout the research process.

It was expected that this review would have identified more examples of linked data research informing decision-making as there have been large government investments into UK data linkage infrastructure, extensive discussions about the benefits for decision-making in early life health, and Scotland has established strong data linkage programmes, such as ScottisH Informatics Programme (SHIP), for over 10 years (Scottish Government 2012a).

The lack of evidence identified in this review could suggest that linked data research is not being used by policymakers. One possible reason for this could be that linked data research is not relevant to the policy and practice agenda or that linked data research has not provided the right solutions that decision-makers are able to implement. Alternatively, the lack of evidence identified in this review could reflect the challenges of measuring research impact (as detailed in section 1.5.5). It could also be interpreted as support for Weiss's theory of enlightenment (1977, 1979, 1982), where research influences policy indirectly through the diffuse of ideas, as it is challenging to identify how the ideas associated with linked data research findings have influenced policy decisions. In addition, it is common for policy changes to occur much later in the research translation timeline, often a long time after the end of the research study. As a result, many changes likely go unreported, making such changes difficult to find in the literature.

The experts consulted as part of this research discussed how the lack of incentives for researchers to trace impact could explain why evidence was not identified in this review. This contradicts the literature presented in section 1.5.3, which implies that demonstration of impact is rewarded by the REF and research funders. In addition, experts consulted in this research highlighted that much of the linked data research has centred around educational outcomes, where exposures and outcomes are measured beyond the age of five years. These studies were beyond the scope of this review, which could also explain why limited evidence was identified.

Further to this, two Information Specialists, Mark Clowes, and Anthea Sutton, produced a report reflecting on their experiences of reviewing the grey literature for a mapping

review that explored how local governments are accessing, linking, and *'using real-world data'* (Clowes and Sutton, 2021). They commented that local authorities *"do not have a well-established tradition of publishing peer reviewed articles",* which was reinforced by a discussion with an expert working in the field of data linkage (see section 2.4.4). This could explain why there was little documentation discovered in this review relating to how local authorities make use of linked data research in decisions related to the early years of life.

Clowes and Sutton (2021) also note that there is much discussion around the sharing of data, but this does not necessarily mean it has been implemented. This creates challenges for identifying instances where data have been shared and used to inform decision-making. Many of the documents identified in this review related to the potential of linked data or detailed the implications of their data linkage findings for policy and practice, but it was unclear whether this had translated into action. Further research is needed to understand if and how linked data research is being used in local and national policy and practice settings. If linked data research is not being used, it is important to understand why it is not being used.

A recommendation for researchers would be to consider the impact of the research early in the research cycle and involve the right stakeholders from the start. In the absence of stakeholder engagement in research, policymakers may be unaware of the research and are, therefore, unable to benefit from its intelligence. Thus, there is a need to bridge this gap between policymakers and researchers, and developing an effective engagement method would be a useful tool for both researchers and policymakers. Further to this, researchers should take responsibility for making the impacts of their research clear and discoverable, and not rely on practitioners and policymakers to document when research has informed their decisions. This would aid the discovery of the evidence sought in this review. In the absence of an impact assessment, it is difficult to ascertain whether linked data research is improving child health outcomes. A recommendation for future research is to develop a way of reporting the impact of linked data research that is discoverable and meaningful for investors, as well as tracking the impact on health outcomes. I acknowledge it will be challenging to isolate the impact of just one research study.

An examination of the identified documents indicated a small knowledge cluster around strategies to promote the use of linked data by decision-makers (n=6), although this did not reveal how successful these strategies are, or best practice around these strategies. The evidence relating to this objective was wide ranging including websites, reports, and peer reviewed literature, each offering a different approach to promoting the use of their findings. The identified strategies ranged from engagement of stakeholders and non-

academic audiences, to disseminating findings in easily accessible formats, to the right audience, and by influential people. This is consistent with the more complex theories of knowledge transfer which recognise the importance of the relationship between policymakers, researchers, and other actors in the knowledge transfer process. The emphasis on engaging the intended users of research is also a key part of the FHI 360 Research Utilisation Framework, which underpins the research conducted in this thesis.

Strategies were often implied rather than explicitly promoting the use of linked data research by decision-makers. For example, documents would describe their engagement activities but not specify the impact that this engagement has on research use by decision-makers. I recognise the challenges of attributing the impact of specific activities on the adoption of research, as detailed in section 1.5.5. However, it is important to understand how researchers can actively encourage the use of their research by policymakers. Further research is needed to understand how best to measure the effectiveness of different strategies for translating research into policy and practice.

Research on the barriers and enablers for policymakers to use research evidence is well established (see Chapter 1). The documents included in this review identified similar barriers to those presented by Oliver *et al.,* (2014). For example, Davis-Kean *et al.,* (2017) found timing and opportunity to be key barriers, aligning with Oliver *et al.,* (2014), although, these barriers were not discussed in detail by Davis-Kean *et al.,* (2017). Hence, there is a gap in knowledge relating to the factors that enable and prevent the use of linked data research by UK decision-makers. It is possible that the barriers and enablers to using linked data research in early years decision-makers are similar to that of other ages or scientific research. More research is needed to understand if these are different from linked data research evidence.

Overall, there was a paucity of documents that demonstrated linked data research influencing UK decision-making around early life health. There are a greater number of studies exploring the strategies to promote the use of linked data research, although, the evidence is still scarce. As the first 1,001 days is a key focus for national policymakers for improving health across the life course, this shows a missed opportunity for linked data to explore issues related to the early years and inform this policy agenda.

2.5.2 Methodological Challenges

The ubiquitous nature of the evidence this review was attempting to identify, meant it was challenging to ensure a comprehensive search of each grey literature source. The websites listed in Table 7 were variable in their usability, where some allowed structured searches while others relied on basic search functions. Hence, specific key words such as *"data linkage"* or *"early years"* were used where complex searches were not applicable. Searching the grey literature meant I was often relocated to other websites, as was also found by information specialists conducting a similar review (Clowes and Sutton, 2021). Like Clowes and Sutton (2021), I also found a number of *"404 File not Found"* messages for documents that were initially thought to be eligible during the title and abstract screening. Hence in some instances, it was not possible to locate the original document identified using the search strategy.

An additional challenge faced during the screening stage was identifying if the research findings related to linked data or children of the right age group. Many birth cohort studies now link observational data to administrative records and are not always clear about whether data linkage has been used to obtain the findings. There were many documents that referred to findings from ALSPAC or GUS, where it was unclear if the findings were a result of the observational cohort data or data linkage to routine records. Consequently, for each reference, I attempted to find the original research study to ascertain whether data linkage was used, and the age of the children involved.

Many of the documents that were found to be ineligible during the full text screening were included at the title and abstract screening stage, as a large proportion of the results were policy documents that did not have an abstract. This is because it is difficult to determine the article's relevance prior to screening the full text, in the absence of an abstract. Eligibility at the title and abstract stage was consequently based on assumptions made about the title e.g., if the title referred to a policy area unrelated to the early years, it was excluded. In addition, many papers describing the use of linked data to investigate child health outcomes were included in the title and abstract screening stage as it was unclear from the abstract whether the research influenced decision-making. This information is often found in the conclusion or discussion. This explains the high volume of documents screened during the full text stage.

It was challenging to determine if a particular webpage was eligible, as often, organisational websites have multiple pages explaining the purpose of their organisation, the potential of their data sources and the populations which they are relevant to, which are separate from the key findings or engagement strategies. For example, ADR UK have one page that details their mission which gives information about data linkage, but details of their public engagement are described on a separate page. Therefore,

decisions on whether to include the page relied on reviewer knowledge of the field of data linkage and exploring other pages of the organisation's website.

2.5.3 Strengths and Limitations

With increased interest in linked data research and its benefits for decision-making, the primary aim of this review was to map the different ways in which linked data research has been used in UK decision-making relating to the early years of life, and to explore the factors affecting the use of linked data as evidence in these decisions. This mapping review followed current best practice with regards to the review methods and was informed by discussions with information specialists and experts in the field. It covers a comprehensive range of sources including both peer-reviewed and grey literature, to maximise the opportunity of capturing relevant evidence. In addition, searching the Overton.ai database increased the likelihood of identifying the influence of linked data research on local decision-making as it collates policy documents from ten UK councils.

The absence of a quality assessment, in accordance with the accepted mapping review procedures, means that this review presents a description of the evidence available on this topic rather than an evaluation of the impact of linked data research. This review focused on publications in English, which is appropriate given that the remit of this research was within the UK.

A possible limitation of this review is that the barriers and facilitators to using linked data research in early years may be no different to using linked data for decision-making in other age groups. Therefore, by focusing on strategies for early years decision-making, this may have omitted key learning from this review. The eligibility criteria for this review could have been extended to include settings beyond the early years, however, this was not possible within the scope of this project as it produced an unmanageable number of results. Future research could look to expand the focus of the mapping review to explore whether there are additional barriers and facilitators to using linked data research in the literature, without limits on the population.

Mapping reviews are useful when there is a large amount of evidence on a topic. Due to the vast amounts of research into linked data, it was anticipated that this review would capture a large amount of evidence. However, the linked data research did not always translate into policy and practice, meaning this systematic map is not as large as expected. There are an increasing number of government interventions utilising linked data, see section 2.5.5, which could be used to populate the map in the future.

Although I deployed a comprehensive set of terms in multiple databases to identify instances where linked data research has informed early years decision-making, I recognise that linked data may have been used but not documented in a published document, especially given my assumption that decision-makers draw on a variety of evidence sources in making decisions.

2.5.4 Implications

Despite the rigorous approach taken to reviewing the literature, it was challenging to identify where linked data research has informed early years decision-making in the UK. This is an interesting finding given the evidence of increased government investment into linked data research and the extensive discussions about the benefits for early years decision-making.

If, in fact, linked data research is being used by decision-makers, this information is not easily discoverable. This has implications for future investment in linked data projects as it is difficult to determine the difference that linked data research has made to UK decision-making and if there is a return on the investment. The responsibility may be with researchers to follow-up and measure the impact of linked data research after it has been completed. This is a key recommendation of this review, although I acknowledge the challenges of tracing the impact of research, especially single research studies.

Policymakers should also document the use of linked data in their decision-making, where possible. Further research is needed to understand if linked data research is being used by policymakers and practitioners, but this is not being reported, or if there are barriers preventing them from utilising linked data research. Primary research could focus on asking those who work on linked data research studies how their findings have been used by local and national decision-makers and asking early years' decision-makers how they have utilised linked data research.

Moreover, further research could explore the barriers and enablers to using linked data research for all ages, rather than limiting the focus to the early years as is the case in this review.

Although many of the documents identified in this review discussed strategies used to promote the use of linked data research, it is unclear what the best practice is for this and how successful these strategies have been. Further research is needed to determine

how to successfully promote the use of linked data research by UK decision-makers in the field of early life health. This is essential given the importance of investing in the early years as a means of improving health across the life course. The strategies that were identified implied that engaging with decision-makers throughout the research cycle for studies utilising data linkage could facilitate the use of these data. Researchers should consider stakeholder engagement as a strategy to promote the use of linked data research by decision-makers.

A reflection from conducting this review is that there is no common language around linked data, which impacts the discoverability of the evidence related to the use of linked data research in decision-making. For example, papers identified by the search strategy discussed the use of routine data, but it was difficult to determine whether multiple routine datasets were linked as part of the research. In addition, many of the research papers that utilised data from studies such as BiB, which links routine data to survey data, were not clear about where the data they used for the research originated from. For example, whether their outcomes were measured in routine data or in survey data. This resulted in these papers being excluded from the review. Therefore, I would recommend a common language around the use of linked routine data for research is developed. This could aid the discovery of evidence sought in this research and allow the benefits of linked data to be recognised.

2.5.5 Papers for future consideration

Two articles that were excluded from this review used linked administrative data to evaluate two government funded initiatives supporting families during the early years and beyond, which demonstrated positive outcomes. One evaluated the Troubled Families programme (Ministry of Housing Communities and Local Government, 2019a) and the other explored child outcomes in relation to the Flying Start programme in Wales (Welsh Government, 2019). They were excluded as a decision regarding the evidence is yet to be taken and the evaluation is still ongoing. However, this is also good evidence that linked data can be used as a large-scale evaluation tool. The evaluations have the potential to be included in the systematic map within the next five years and it is postulated they will have similar impacts to the Family Nurse Partnership as the linked data are being used to evaluate government funded initiatives.

2.6 Chapter Summary

The overall aim of this mapping review was to improve transparency around how linked data research is being used beyond academia to improve outcomes for children and families. Despite many documents describing the benefits of linking routine datasets to other sources and using them as an evidence base for policy and practice, limited evidence of linked data informing early years decision-making was identified in this review. Therefore, this review demonstrates a gap in the knowledge for whether linked data research can inform early years' decision-making and the factors influencing its use. This creates challenges when evaluating the benefits of linked data research for influencing policy, which are used to justify future investments.

It is important to understand how linked data have previously been used in decision-making as this can inform how researchers using these data can engage in activities that promote the use of their research findings. There is evidence in both the research-policy gap literature and in studies identified in this review, that engagement of policymakers in research can support research utilisation. This could suggest that engaging early years policymakers in linked data research could facilitate the translation of linked data research into decision-making. This finding, alongside the literature detailed in Chapter 1, informed Chapter 4 of this thesis.

This systematic map is the starting point for collating information about how linked data research have influenced UK decision-making related to early life health. Increasing investment into data linkage for early years research means that this map can be added to overtime as the investment begins to produce outputs, including the Flying Start linked data evaluation project and Troubled Families evaluation, identified in section 2.5.5.

Overall, the size of the discoverable evidence base is modest, and the quality of the evidence is limited, given that many of the included documents only had a small section relevant to the review objectives. Therefore, it is unclear how linked data research are being used by decision-makers. This could suggest a missed opportunity for linked data research to focus on the early years of life and appeal to the first 1,001 days of life policy agenda. The research in this thesis aims to address the gaps in the knowledge by exploring whether linked data research can be used as a local health intelligence tool for child and maternal health. This review also identified a gap in the knowledge related to the barriers and enablers for early years' decision-makers to use linked data research. This informed the direction of the primary research detailed in Chapter 6, where I sought to understand how local decision-makers could be engaged and supported to use linked data research.

Section B: Case Study

This section contains the methods and findings of the case study exploration across 4 chapters.

- Chapter 3 describes the research context. This includes a description of the Born and Bred in (BaBi) Network, which is the case study explored in this thesis. It presents the BaBi Local Health Intelligence (LHI) model that underpins the aims of the BaBi Network and explains how this is applied to this research. This chapter also sets out the aims and objectives of this thesis and justifies the methods used to address these.
- Chapter 4 presents the methods, outputs, and discussion for the first stage of the BaBi LHI model: identifying research priorities for linked data research around the theme of early life health.
- Chapter 5 presents the methods, findings, and discussion for the second stage of the BaBi LHI model: addressing local research priorities using linked data.
- Chapter 6 presents the methods, findings and discussion for a qualitative study exploring how local early years decision-makers can be engaged and supported to make use of linked data research.

Chapter 3: Research Context and Methodology

3.1 Introduction

Chapter 1 of this thesis established the potential of using linked routine data to inform decisions around early life health, specifically, decisions at the local level. This is important because the introduction of the ICS has created more opportunities for local data driven decision-making. The idea that high quality data should be used to inform health care decision-making has led to the development of models and frameworks to explain how this could work (see section 1.4).

However, Chapter 2 found limited evidence of linked routine data being used to inform decisions around early life health and/or successful strategies for promoting the use of linked data research findings. Therefore, it is unclear whether these data are being used by decision-makers but that this is not well documented, or if there are barriers to using linked data research to inform decision-making. It is important that we understand if there are challenges associated with using linked data to inform decision-making.

This chapter describes a series of studies that are being established across the UK, with the intention of linking routine data for early years research. These studies form the Born and Bred in (BaBi) Network and are being coordinated by a research team at BiB. I use the BaBi Network as a case study to explore whether linked routine data can be used to support local early years decision-making and to identify the challenges of using these data in this way. The aim is to address the knowledge gaps described in Chapter 2 and to provide guidance for teams setting up studies as part of the BaBi Network. I draw on the ideas of learning health systems, data driven decision-making and research utilisation (see Chapter 1) as these underpin the aims of the BaBi Network. This chapter details the research setting and the context behind the methods chosen to address the thesis aims and objectives. It also outlines the assumptions I made in designing this research.

This chapter begins with an introduction to BiB and the development of the BaBi Network. I then discuss the model that underpins the aims of the BaBi Network, and the research conducted in this thesis. This is followed by the thesis aim and objectives. Section 3.5 justifies why a case study approach is appropriate for this research and section 3.6 sets out the ontological and epistemological assumptions that informed this research. Section 3.7 provides a summary of how the Covid-19 pandemic impacted this research, which is followed by the concluding section.

At the time of completing this research, the BaBi Network, and the local BaBi study used in this research (Born in Bradford 4 All), did not have a published protocol. Hence, the information detailed in this section was informed by documents that were submitted as part of the ethical approval process and by conversations with the BaBi coordinating centre team.

3.2 Born in Bradford

The city of Bradford, in the North of England, is one of the most deprived areas of the UK and has many associated public health problems. This includes higher than average infant mortality rates (6.09 per 1000 compared to 3.95 per 1000 in England), above average obesity in pregnancy rates (24.1%, 22.1% in Bradford and the rest of England respectively) and a higher rate of smoking in pregnancy (16.5%) compared with the rest of England (12.8%) (Public Health England, 2021a; Raynor *et al.,* 2008). In addition, infant mortality is highest for babies of Pakistani origin, who account for almost half the babies born in Bradford (Raynor et al., 2008). To tackle these issues, it is important to understand the complex nature of the population facing these issues, in the context of the wider social, economic, and environmental determinants that shape children's health.

Consequently, in 2007, the BiB longitudinal birth cohort study was established to examine the impact of genetic, nutritional, environmental, behavioural, and social factors on the health and development of the population of Bradford (Wright *et al.,* 2013). It recruited over 13,500 children who were born at Bradford Royal Infirmary between March 2007 and December 2010 and their parents (Wright *et al.,* 2013; Raynor *et al.,* 2008). Upon recruitment, pregnant women were weighed, measured, and biological samples were acquired and stored. Detailed information was obtained in the form of a questionnaire and permission to link data that have been routinely collected about them was given (Wright *et al.,* 2013).

BiB differentiates itself from other birth cohort studies as much of BiB's work is underpinned by meaningful community engagement and involvement. The BiB cohort was designed to work in partnership with local services, to provide evidence that is useful to policymakers and practitioners (Wright *et al.,* 2021). BiB also work closely with local families to set research priorities, which ensures the outputs are relevant to the real issues faced by the local population. This ethos of community and stakeholder engagement has influenced how the research was conducted in this thesis.

BiB acknowledge the importance of place-based research for addressing local needs. BiB represents families from over 40 different countries and around 45% of the BiB children are of Pakistani heritage and 40% are of a White British heritage. Hence, BiB research can allow for more tailored interventions to suit the needs of this population (Wright *et al.,* 2021).

BiB has been successful in achieving impact through influencing policies and practice. For example, using the routine data available from the BiB cohort, researchers were able to identify that poor maternal mental health increases the risk of poor child mental health at age three years. They showed that although Pakistani women are more at risk of mental ill health, they were only half as likely to have a recorded diagnosis in their primary care record compared with White British women (Born in Bradford, 2019). Following these findings, the data capture systems used by midwives and health visitors were improved, allowing more women to be identified as needing support and there was a focus on improving the pathways to treatment for ethnic minority groups. This example highlights how important linked routine data can be for research and decision-making.

However, this example was not included in the mapping review in Chapter 2 as it was unclear in the document reporting these findings, that they were generated using linked routine data. Born in Bradford (2019) use the term *"current data systems"* and do not link the policy impact to the original research study. This meant reviewers were unable to determine whether this research was produced using linked routine data. Following the completion of the mapping review, I spoke with researchers at BiB and they were able to confirm that this research was produced using linked routine data. This further supports the recommendation to develop a common language for linked data research, to increase the transparency around how linked data research has informed decision-making.

3.2.1 Born in Bradford 4 All

In 2019, the BiB research team launched their new data linkage cohort study, Born in Bradford 4 All (BiB4All) (Bradford Teaching Hospitals NHS Foundation Trust, 2021a). BiB4All gains opt-in consent from women, during pregnancy or after giving birth, to access and use data that are routinely collected about themselves and their child for research purposes. Every woman who is booked to receive maternity care at Bradford Teaching Hospital NHS Foundation Trust (BTHFT) is invited to join the study by a trained midwife during a routine appointment. Consent is then recorded in the electronic patient record, providing an efficient, paperless process for recruitment, that is not reliant on research team capacity. The legal basis for using routine data for research, as part of the

BiB4All study, can be found on their website (https://borninbradford.nhs.uk/privacy-policy/).

BiB4All was established to harness the power of routine data on a wider scale after its successful and extensive use in the original BiB cohort study (Wright et al., 2013). This provides the opportunity to:

- Better describe child health and development in Bradford and identify whether patterns exist in the data that might help in the early identification of families with poor health and development.
- Evaluate the impact of a wide range of interventions or policy changes.

Figure 7 depicts some of the organisations that can be contacted for information on BiB4All participants, which can then be linked together for research.

Figure 7 Sources of routine data that can be linked for BiB4All participants *

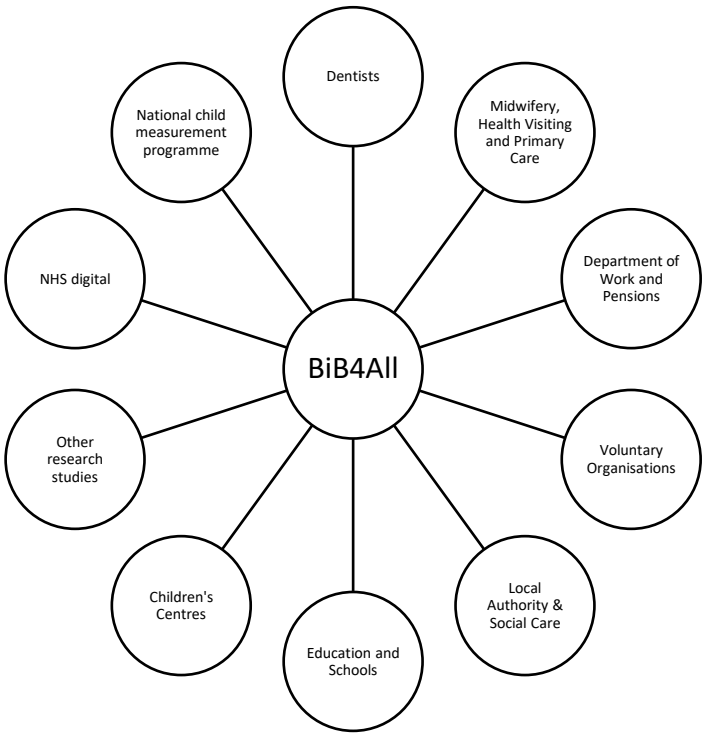


**Figure adapted from 'BiB4All' by Born in Bradford (2023).*

Hence, once routine data are connected, it can be a powerful tool used to answer locally relevant research questions.

In addition to linkage of routinely collected data, consent includes permission for future contact for research. I refer to this as the *'consent to contact'* process in this thesis. This

provides a source of potential participants for specific research studies that are likely to benefit that population.

Proof of concept has been established in Bradford, where more than 15,500 women and their babies have joined the cohort since 2019. The nature of embedding consent into routine practice means that study recruitment in Bradford continued during Covid-19 lock-down, unlike many other research cohorts.

The stakeholders likely to benefit from research produced using these data, at the time of completing this research, included midwives, health visiting teams, and early years services as they are contributing their data, as well as commissioners and policymakers. Hence, these stakeholders are the focus of this research.

The representativeness of this cohort, with respect to the population of Bradford, is explored in Chapter 5.

3.2.1.1 BiB4All Data Linkage Procedures

This section summarises how data are linked for BiB4All participants.

Data sharing agreements are established between BiB and the relevant data custodians for each data source linked as part of BiB4All. These data sharing agreements detail the method of secure transfer of data to BiB and the frequency of data updates to be shared. Each agreement is different based on the need of the data source (Bradford Institute for Health Research, 2020, Unpublished).

There is no central procurement arrangement for an electronic patient record system across health and care providers. This could explain why linkage across records for individuals and families is not commonplace. GPs and health visitors in Bradford record their data on the electronic health record system, Systm One. Midwives in Bradford operate a different electronic health record system, as well as noting down information on paper medical records (Bradford Teaching Hospitals NHS Foundation Trust, 2023a).

To link this information together, a list of consented participants BiB4All study ID numbers are combined with identifiers such as name, date of birth, address, and/or NHS number. These are shared with the data source via a secure data transfer method. The data source then attaches the relevant routine data to the individuals and returns the data via a secure data transfer method. Participant identifiers are removed before the data are returned, leaving only the unique BiB4All study ID as the link. A new version of this list of

BiB4All IDs is created at the time of the linkage, to ensure any participants that have withdrawn are not included. Where feasible and appropriate, OpenPseudonymiser may be used to create a pseudonymous key for data linkage and matching purposes (Bradford Institute for Health Research, 2020, Unpublished).

To link to primary care records, extracts of Systm One records are matched to cohort participants' records based on the deterministic requirement that each of the four fields match (NHS number, surname, date of birth, gender) by the data provider. If unique identifiers are not available, iterative deterministic matching on the basis of multiple sets of non-unique identifiers is used. The BiB research team have these data sharing agreements in place and receive regular linked data extracts from the provider.

More details on data capture, storage, and access management can be found in the '*Born in Bradford: A data linkage cohort study of babies born in Bradford and their mothers: Protocol. Version 4.0*', which is available on request (Bradford Institute for Health Research, 2020, Unpublished).

Currently, the process for accessing these data is to apply to the BiB Executive Committee who approve the research study based on scientific merit and ability to deliver the request.

3.2.2 Born and Bred in Network

In addition to maximising the use of routine data in Bradford, the BiB4All team are supporting the development of a network of local data linkage cohorts across the UK, known as the BaBi Network. Each site intends to set up its own data linkage study in a similar way to BiB4All. This will allow each area to explore locally relevant questions, using the linked data as a '*local health intelligence tool'* for child and maternal health.

In the context of this thesis, the term *'local health intelligence'* is defined as information, data, knowledge, and evidence that results in local, evidence-based action for a defined local population (World Health Organisation, 2014). Thus, the term *'local health intelligence tool'* is used here to describe the process of using linked data research to inform the provision of local services to better meet the needs of their local population. I discuss these ideas further in the next section.

The BaBi Network began with developing local data linkage cohorts in five local areas which are referred to as the BaBi pilot sites. This was important for establishing key

processes before expanding further. These pilot sites included the original site (BiB4All), BaBi Leeds, BaBi Wakefield, BaBi Doncaster and BaBi East London. The BiB4All cohort has retained its title locally as it was important that it was seen as part of the BiB family of projects, which have an established trusted reputation within the community. Outside of Bradford, the BiB4All cohort is often referred to as *'BaBi Bradford'*. Following the success of the BaBi pilot sites, other local areas were interested in joining the Network. For an up-to-date list of the sites currently participating in the BaBi Network see https://www.babinetwork.co.uk.

As part of the BaBi Network model, the BiB4All team act as the BaBi Network Coordinating Centre, where they provide strategic research support to each BaBi site, to replicate what has been successful in Bradford. They monitor each site to ensure they have the correct processes, documentation, and procedures in place, as well as provide governance. They are also the sponsor of the research. Information about the BaBi Network Coordinating Centre team can be found at https://www.babinetwork.co.uk.

Each BaBi site is the Data Controller for their BaBi cohort and can make local decisions on how their data are used.

The BaBi Network also aims to develop a meta-cohort, bringing together data from across each of the BaBi sites to address common questions and those of national relevance. Hence, each BaBi site agreed to provide regular, anonymised extracts of linked data to the central BaBi Network Co-ordinating Centre, who will then collate these extracts to produce an anonymised meta-cohort database that can be analysed. The current plans are for the BaBi Network Co-ordinating Centre to be the Data Controller for the meta-cohort.

Figure 8 demonstrates the structure of the BaBi Network with the BaBi pilot sites.

Figure 8 Structure of the BaBi Network

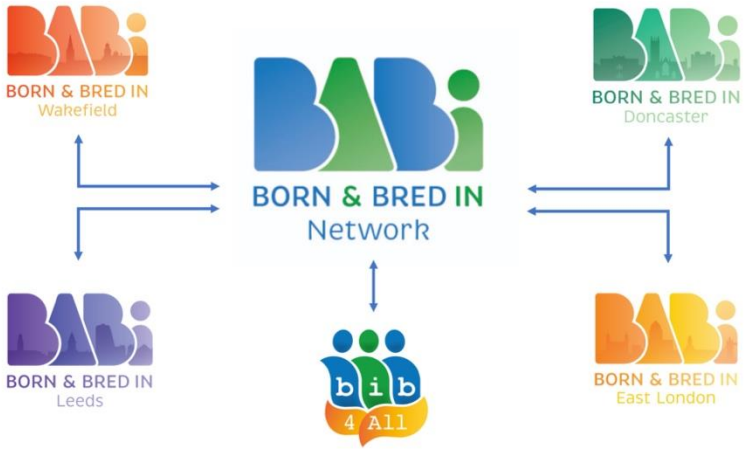


**Figure adapted from 'Born and Bred in (BaBi)' by Bradford Teaching Hospitals NHS Foundation Trust (2021b).*

As BaBi is a collaboration between all BaBi sites, there is a central repository on Google Drive, which is accessible to everyone that works on BaBi. This allows people to add resources and share documents and updates.

This model of data linkage allows for a cohort of people for every generation to support local families. This investment in infrastructure will allow broad ranging data to be used by multiple local partners, creating the potential for every place to establish a sustainable maternal and child health learning system, which promotes genuine engagement from the public and health and care professionals. The BaBi Network provides the opportunity to learn more about the power of routine data and its limitations in its current form.

3.2.2.1 BaBi Network Key Meetings

As part of the BaBi Network model, there are a range of groups that meet regularly that are important to the development of the BaBi Network. These are summarised in Table 16. I attended these meetings throughout my PhD, and I utilised these networks when conducting the research detailed in this thesis.

A key part of the BaBi model is collaboration with local services, which aligns with the ICS model. Therefore, the BaBi steering groups play a crucial role in the success of these studies.

Table 16 BaBi Network Meetings

Group	Purpose of the Group	Members of the Group/Attendees at the meeting	Frequency of the Meeting
Local Management Groups	Each local BaBi pilot site has a management group that work closely together to ensure the implementation of the cohort at their site.	▪ A principal investigator to drive the project locally. ▪ A BaBi Project Lead to oversee and complete set up; to represent the local site at meetings; create links and partnerships with local organisations; identify and apply for funding to support their site; work with academic partners to identify research opportunities and to disseminate the findings to stakeholders. ▪ A BaBi research midwife to train clinical staff to seek consent, monitor recruitment and troubleshoot any barriers, ensure all research governance is adhered to and to promote the study. ▪ Data/IT support to embed the programme within the local maternity electronic patient record, create and manage the local BaBi database and link the data at the site level.	Monthly
Local Steering/Partnership Groups	Each of the pilot sites engage key partners to give advice and/or make decisions about the development of their cohort.	Local authority representatives, academic institutions, local services, clinical leaders, service users' and education services.	Approximately every four to six weeks
National BaBi Network Steering group	To strategically lead the network, agree meta-cohort data releases, support sustainability and development.	The BaBi Network Academic Director, Bradford Project Investigator and National Speciality Leads in child health, maternity, and data (invite only).	Two or three times a year
BaBi Network Management Meeting	These meetings took place during the set-up of BaBi pilot sites. These meetings proved invaluable for problem solving, decision-making and sharing good practice that allowed the network to progress. They also provided an effective channel of communication between the sites.	▪ BaBi Network Coordinating Centre ▪ Key members of the local BaBi pilot sites	Every six weeks
BaBi Network Coordinating Centre Team Meeting	To discuss and approve data access requests as well as requests from local areas to be part of the BaBi Network.		Monthly

Once BaBi sites were up and running, the frequency and purpose of the BaBi Network Management meeting changed. They were replaced with a meeting attended by the BaBi Network Coordinating Centre and BaBi site Principal Investigators, and multiple shorter meetings. These shorter meetings were set up with the intention of bringing together people who were carrying out similar roles across the BaBi sites to provide peer support and continue to share good practice. For example, there is meeting for the BaBi research midwives, which occurs twice a month, and a monthly BaBi data operations meeting.

3.2.2.2 BaBi Local Health Intelligence Model

Wright *et al.*, (2021) describes how the goal is for the BaBi studies to be used locally as a learning health system for maternal and child health. See section 1.4.3 for more information about learning health systems.

Chapter 1 of this thesis presents the stages of using data to drive decision-making (Figure 4) as set out by the United States Department of Health and Human Services. It involves four stages and is cyclical in nature: (1) Formulating key questions; (2) Collecting and analysing relevant data; (3) Communicating results to relevant decision-makers and (4) Making use of these data to change processes, organisations, or systems.

Two directors of the BaBi Network, Maria Bryant and Sally Bridges, developed a model describing how data from the BaBi studies could be used to drive local decision-making (Bryant and Bridges, 2021, Unpublished). This is known as the BaBi Local Health Intelligence (LHI) Model, which is presented in Figure 9. This model is underpinned by the principles of data driven decision-making and learning health systems.

Figure 9 BaBi Local Health Intelligence Model*

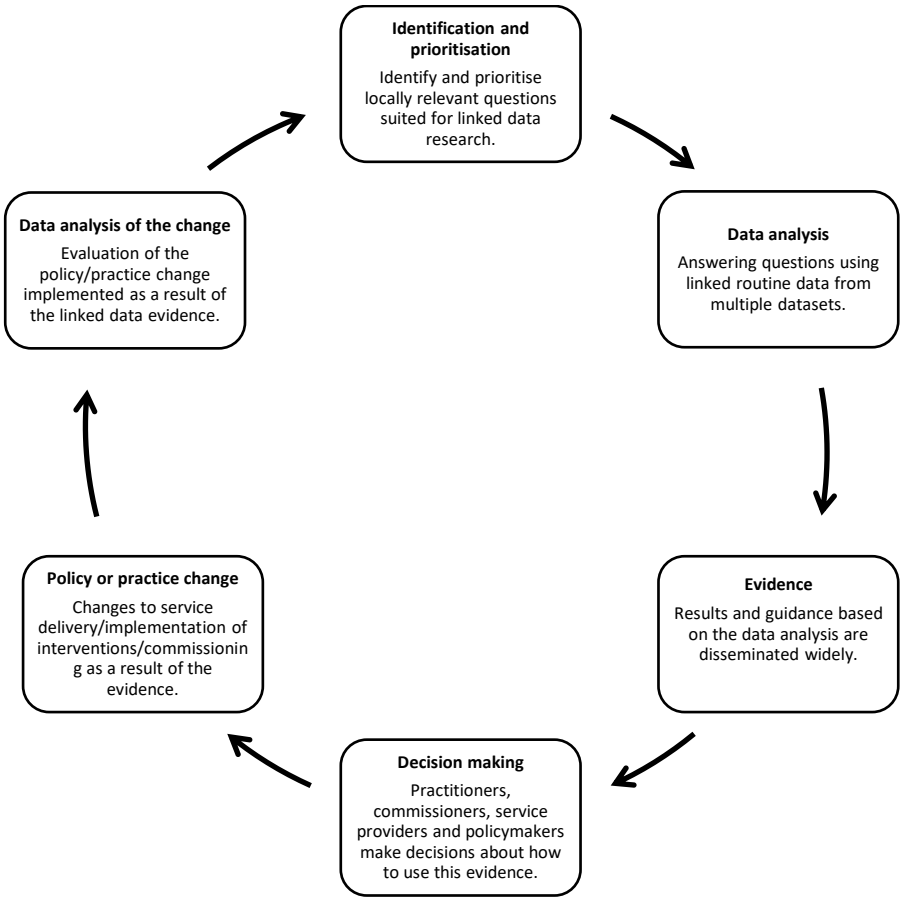


**Figure adapted from (Bryant and Bridges, 2021, Unpublished)*

The stages of this model also align with some of the key phases of the FHI 360 Research Utilisation Framework presented in Figure 5, section 1.4.3 (Kim *et al.,* 2018). As such, the first stage of the BaBi LHI model involves identifying and prioritising locally relevant research questions that can be addressed using linked routine data. This corresponds to the foundational phase of the FHI 360 Research Utilisation Framework; hence it is important to engage stakeholders at this stage to understand their needs. The next stage of the BaBi LHI model involves bringing together routine data from multiple sources to address the identified research priorities. This relates to the research phase of the FHI 360 Research Utilisation Framework. The data analysis stage then produces some results, from which recommendations for policy and practice can be made. These results and recommendations are then disseminated widely to relevant stakeholders and used to inform decision-making. This can result in changes being made to policy and practice. Thus, the evidence, decision-making, and policy and practice stages of the BaBi LHI model form part of the translational phase of the FHI 360 Research Utilisation Framework, where research findings are turned into actionable products that are widely disseminated and used. Finally, the last stage of the BaBi LHI model involves evaluating

the changes made to policy and practice change. This aligns with the activities conducted as part of the institutionalisation phase of the FHI 360 Research Utilisation Framework.

As with the FHI 360 Research Utilisation Framework, the BaBi LHI model allows for continuous learning and feedback loops throughout the cycle. Box 8 provides an example from a previous study that used routine data from a BiB cohort, to illustrate how the BaBi LHI process could work.

Box 8 Example of how linked routine data can be used to inform decision-making

Pettinger *et al.,* (2020) investigated the relationship between prematurity at birth and school readiness using routine data from multiple sources. As a result of this research, new interventions are being developed to ensure that children who need support when starting school receive it.

The BaBi studies could enable a local area to evaluate such an intervention without the need to establish a new study specifically for that. Data from the BaBi studies could also be used to answer questions about uptake of services, including identifying groups in the local population who are more or less engaged with a service or intervention. The consent of participants in a BaBi study includes being able to follow up with participants to be involved with other research projects. This means that more detailed research could be carried out to further understand what works to improve outcomes.

As discussed in section 1.4.2, the WYHCP is an ICS that supports over 2.4 million people and one of their main priority areas is to improve outcomes for children and families (West Yorkshire Health and Care Partnership, 2023). Three of the five local areas covered by this ICS (Bradford, Leeds and Wakefield) are part of the BaBi Network. Thus, data from these local BaBi studies have the potential to inform the WYHCP decision-making.

As the BaBi Network is newly established, there is little understanding of how each of these stages might work or how these data can be used to drive decision-making. Chapter 1 discussed how linked data research is complex and may face additional challenges to other scientific research. Chapter 2 found limited evidence regarding how linked routine data has been used to inform early years decision-making. Hence, this thesis will use the BaBi LHI model as a basis to explore how linked data research can be used to support early years decision-making in the UK.

When exploring the application of the BaBi LHI model as part of this thesis, it will be important to consider the activities and actors depicted at each of the stages of the FHI 360 Research Utilisation Framework, to help guide this research utilisation process. The FHI 360 Research Utilisation Framework recognises the complexities of translating research into practice, unlike the linear models of knowledge translation, and draws on some of the ideas from political theories (see section 1.5). Decision-makers likely make use of a range of evidence sources when making decisions, as discussed in Chapter 1. Therefore, this research considers how linked routine data can be used as one of these sources, to allow for better informed early life health decision-making. This aligns with more complex theories of knowledge transfer such as the *'interactive model'* and *'enlightenment'* theory (Weiss, 1977, 1979, 1982).

3.2.2.3 Local Data Accelerator funding

In 2021, the Department of Levelling up, Housing and Communities, formerly the Ministry of Housing and Local Government, made £7.9m of funding available to support children and families through better use of local data (Ministry of Housing, Communities and Local Government, 2021). This was known as the Local Data Accelerator fund.

BaBi teams in Leeds, Doncaster, and Wakefield, as well as the BaBi Network Coordinating Centre collaborated and were awarded funding as part of this bid. The research I conducted as part of this thesis informed the funding application for this bid and contributed to the bid objectives. The four specific objectives of this bid were:

1) To establish data linkage processes to make best use of consented connected data at a local level.
2) To establish a network of electronic birth cohort (e-cohort) studies to develop skills and methods for linking and using data to inform policy and practice. Where electronic birth cohort refers to the data linkage model adopted by the BaBi studies.
3) Conduct local level prioritisation activities to produce a plan to use the consented data over time.
4) Develop a toolkit to support other sites to set up similar connected cohorts.

The proposed toolkit aimed to include a step- by- step guide to developing a data linkage study, template data sharing and collaboration agreements, information governance advice, and engagement materials for communities and professionals. It was intended to be used by the new BaBi management teams to enable set up and participation as a BaBi site. This may be referred to in this thesis as the 'BaBi Toolkit'.

Each local BaBi area has its own unique geography, demographics, established links and connections so this toolkit aimed to provide a flexible, principles-based guide that can be adapted to suit the needs to local stakeholders and families. The toolkit is important as BaBi is quite different to traditional research studies that NHS sites are familiar with, and setting up a BaBi cohort is a long-term investment in improving the health and wellbeing of that local community.

3.3 Thesis Aim

This thesis aims to understand how linked data can be used as a local health intelligence tool for child and maternal health, within the context of the BaBi studies. The overarching aim is to generate understanding about how researchers and research users (such as practitioners, commissioners, and service providers) can be supported throughout this process.

I will produce practical recommendations for new and existing BaBi sites on how to engage and support the use of linked data in their local areas, at each stage of the BaBi LHI model. These recommendations may also support others using linked routinely collected data for research, outside of the BaBi Network. This is important as there is little evidence on how linked data can inform early years policy and practice in the literature.

3.4 Thesis Objectives

The objectives of this thesis are directly informed by the stages of the BaBi LHI model and the gaps in the knowledge identified in Chapters 1 and 2. I will focus on the first four stages of the BaBi LHI model (identification and prioritisation, data analysis, evidence, and decision-making). This can inform future research which will focus on evaluating the changes made.

The key objectives of this thesis are:

1) To identify research priorities around child and maternal health to be addressed using BaBi data (stage one of the BaBi LHI model).
2) To explore whether local research priorities can be addressed using linked routine data from the BiB4All study (stage two of the BaBi LHI model).

3) To understand the perspectives of local decision-makers towards evidence produced using linked routine data and using it as a local health intelligence tool for child and maternal health.
4) To identify the support needs of local decision-makers to use evidence produced using linked routine data as a local health intelligence tool for child and maternal health (relates to stages three and four of the BaBi LHI model).

Chapter 4 addresses research objective one, Chapter 5 addresses research objective two and objectives three and four will be addressed in Chapter 6. These chapters outline the background, methods, and results for each of these studies and report on any challenges faced. Therefore, this thesis follows a journal-style as each of the studies correspond to a stage of the BaBi LHI model and the methods of each study develop from the results of the previous study.

Objective one directly informed the BaBi toolkit as part of the Local Data Accelerator Fund objectives. Objectives three and four were directly informed by the gaps in the knowledge identified in Chapter 2.

3.5 Case Study Approach

This project adopts a case study approach. A case study approach is an established research design that is extensively used in the field of social sciences. There are many ways to define a case study, however, the central tenet is exploring an event or phenomena in-depth, in its natural context. This contrasts to an *'experimental design'* such as a randomised control trial, where the investigator exerts control over the variables of interest (Crowe *et al.,* 2011). Hence, a case study approach can allow for an in-depth, multifaceted enquiry of complex issues in their real-life settings to provide broader lessons, for example, to explore experiences of a new policy initiative or service development (Crowe *et al.,* 2011).

The crucial stages of undertaking a case study are defining the case; selecting the case(s); collecting and analysing the data; interpreting the data and reporting the findings. In this project, the case is the BaBi Network as described above. These are local areas where local data linkage cohorts are being established. The *'problem'* is lack of understanding around how linked data research can inform local early years decision-making.

Thus, in the context of this thesis, a case study approach allows for an exploration into how linked data research can be used in local areas setting up BaBi studies. This can provide broader lessons on how to support the use of linked data research in local decision-making.

A case can be considered as a bounded system that exists independently of the inquiry. Each case should have a pre-defined boundary which clarifies what is covered by the case study such as the time period, the relevant social group, or geographical area. It is important to respect these boundaries as we come to understand how the people operating within this case view their world (Stake, 1978). Therefore, my research considers the BiB4All, BaBi Leeds, BaBi Doncaster, BaBi Wakefield and BaBi East London studies, as these were the BaBi pilot sites at the time of the completing this research.

Much of this research focuses specifically on the BiB4All cohort as this study was the most developed at the time of completing this research. However, key aspects of the BaBi LHI model were also explored for other local BaBi sites to generate a broader appreciation of the issue. This is known as a collective case study as it involves studying multiple cases simultaneously or sequentially. As BaBi East London was established after the other BaBi pilot sites, I do not explore this as a case study in some parts of this thesis. Thus, it is also an instrumental case study as I am studying particular BaBi sites in Yorkshire as an exemplar of the experiences of researchers and decision-makers more generally. Instrumental case studies seek to understand an issue in a particular population to generate a number of findings that can be transferable to other contexts (Stake, 1995; Crowe *et al.,* 2011).

By exploring the BaBi LHI model in multiple sites, it allowed the methods to be tested in a number of different contexts. For example, NHS trusts in each of the cases have differing levels of research experience and, therefore, it was important to understand these differences.

As part of my PhD, I was embedded in the BiB4All research team. This allowed me to develop a network of contacts within the BaBi Network that I could utilise as part of this research. The sites involved in my research have agreed to be part of the BaBi pilot and were, therefore, willing to be a part of my research.

3.5.1 Criticisms of Case Study Approaches

There are a number of criticisms of case study approaches such as lack of scientific rigour and generalisability (Crowe *et al.,* 2011; Gomm *et al.,* 2009). I focus on these two criticisms as these are most relevant to the research conducted in this thesis.

The generalisability of case study research has been the subject of ongoing debate. As discussed, the aim of some case study approaches is to draw conclusions about some general phenomena or about the wider population of cases. Lincoln and Guba (2009) identify a number of problems with the idea that generalisation is the aim of science. They argue that researchers are not faced with the choice between searching for general laws or studying unique cases, but that research can be somewhere in between. They suggest that conclusions in one study might hold in another context and that case studies can produce *'working hypotheses'* that can be explored to understand other cases (Gomm *et al.,* 2009). For this to happen, researchers must provide *'thick descriptions'* of the cases they are studying.

Donmoyer (2009) argues that adopting generalisability in case study approaches is not appropriate. Donmoyer (2009) criticises Lincoln and Guba (2009) for assuming that we can only use knowledge from one case study to understand another if the two cases are deemed similar. Instead, Donmoyer argues that differences between cases can be illuminating and that case studies may facilitate learning by substituting for first-hand experience.

Schofield *et al.,* (2009) insists that case study researchers can put forward overall conclusions as long as they consider what they want to generalise and design their study to maximise the generalisability of their findings. Gomme *et al.,* (2009) also claim that case study researchers can make general conclusions but underline the danger of drawing misleading conclusions about aggregates from a few study cases.

This thesis takes the perspective of a widely cited paper by Stake (1978). Stake (1978) argues that case studies can have general relevance even though they may fail to provide a sound basis for scientific generalisation in a conventional sense. If research is to be valuable, it needs to be framed in the same terms as the everyday experiences through which we learn about the world. Thus, a strength of case study design is that they can provide knowledge of experiences in the form of full and thorough knowledge. Stake (1978) concludes that case study researchers do not need to provide generalisations but describe the case they have studied properly, in a way that captures its unique features.

Case study approaches offer important advantages over more conventional kinds of research as they can provide personal perspectives as well as an in-depth exploration in a real-world setting. This is appropriate for addressing the aim of this thesis.

To address these concerns over scientific rigour, I have been transparent throughout my research process about the choices I have made and how my involvement has influenced the data collected and interpretation. The methods developed in this project will be applied by additional BaBi sites beyond the completion of my PhD and continue to be evaluated.

3.6 Epistemology and Ontology

Case studies may be approached in different ways depending on the ontological and epistemological viewpoint of the researcher. Ontology refers to beliefs about the nature of reality and epistemology refers to beliefs about knowledge and how its acquired (Al-Saadi, 2014).

I approached this research from a critical realist perspective, meaning reality is believed to exist independently of those who observe it, and that this reality is only accessible through individuals' perceptions (Bhaskar, 1975). It denies that we can have any *'objective'* knowledge of the world and accepts the possibility of alternative accounts of a phenomenon that can be equally valid (Maxwell, 2012). It also recognises the way that individuals give meaning to their experiences and how these meanings are influenced by the wider political and social context (Braun and Clarke, 2006).

One of the central tenets of critical realism is that ontology is not reducible to epistemology. As such, critical realism suggests there are three levels of reality: the empirical, the actual, and the real. The empirical level refers to the event as we experience and observe it, however, how we experience an event is always mediated through the filter of human interpretation. The actual level describes how events occur whether humans experience or interpret them and these true occurrences often differ from what is observed at the empirical level. The real level is where causal mechanisms exist, which are the underlying generative mechanisms driving the actual and the empirical (Fletcher, 2017). Fletcher (2017) suggests that causal mechanisms are social products that can be understood through phenomena at the empirical level, making phenomena relevant for scientific exploration.

Hence, a critical realist approach implies that there are layers of reality and different forms of knowledge, some which can be knowable and some which are inferred by the researchers.

Critical realists use *'retroductive'* reasoning, where larger social problems (reality) are explored through rational judgements, theories and causal mechanisms that may impact the levels of reality (Fletcher, 2017). The work conducted in this thesis takes a critical realist perspective to consider the problem (e.g., limited use of linked data research to support early years decision-making) and the barriers and facilitators that could explain the problem, using the BaBi LHI model as a guide.

As a result, the findings from this research project are embedded in the context in which they were collected and analysed. Throughout this project, I have reflected on and been transparent about how my beliefs and assumptions have likely impacted the research outputs. I have also clearly outlined the research setting for each of the research studies presented in this thesis, which address the different stages of the BaBi LHI model.

I approached this research with assumptions about knowledge translation, which are detailed in Chapter 1. These assumptions influenced how I interpreted the BaBi LHI model and subsequently the research questions addressed. These assumptions also influenced my interpretation of the findings and the subsequent recommendations for BaBi sites. Being embedded within the BaBi Network has also influenced the design of this research and my interpretation of the research outputs.

Critical realism does not favour a particular methodology (i.e., qualitative, or quantitative), instead the choice of research method depends on using the best tools to uncover the knowledge sought (Zachariadis et al., 2013). Thus, critical realism was well suited to this research as there are varying explorations within the BaBi LHI model.

3.7 Covid-19

This research was conducted during the Covid-19 pandemic. This meant that face-to-face research was not always possible, and this influenced the methods chosen for this research. The Covid-19 pandemic resulted in unprecedented pressures on public and third sector organisations. This research relied on engagement from health and care professionals. This created challenges in conducting the research during this time, especially as there were periods where NHS trusts paused engagement with research to prioritise clinical care. This resulted in large gaps between carrying out engagement

activities. To enhance accessibility of this research for a broad range of stakeholders, I undertook most of the data collection and engagement for this research online.

3.8 Chapter Summary

This chapter introduced the BaBi Network, a series of studies that gain consent from pregnant women to access and use data that are routinely collected about themselves and their child for research purposes. The BaBi LHI model, presented in section 3.2.2.2, describes how data from the BaBi studies could be used to drive local decision-making. This model is underpinned by theories of learning health systems and data driven decision-making and provides a lens for exploring how linked data research can be used to support early years decision-making in the UK.

The research detailed in this thesis uses the BaBi Network as a case study to explore how linked routine data can be used as a local health intelligence tool for child and maternal health. The aim is to understand how researchers and research users can be supported to use the BaBi data in this way and to identify any challenges that would prevent the use of linked routine data in decision-making. As the focus is on the utilisation of linked data research by decision-makers, the literature on research impact and research utilisation, described in section 1.5, informed the research conducted in this thesis. Specifically, I consider the ideas presented in the FHI 360 Research Utilisation Framework on the activities and important actors involved at each stage of the BaBi LHI model.

Chapters 4-6 present three studies that address the research objectives set out in section 3.4. Chapter 4 reports on the first phase of this research project which involves identifying priority areas of child and maternal health with health professionals, commissioners, researchers, and members of the public, for research. Chapter 5 explores whether these priority areas can be addressed using data from the BiB4All study. Chapter 6 explores the perspectives of local early years decision-makers, in areas setting up BaBi studies, towards linked data research and how they can be engaged and supported to make use of research produced with linked routine data. Each chapter informs the research conducted in the subsequent chapters.

By understanding the context in which the BaBi studies are being established, it can help determine how linked data research from those studies can be implemented into decision-making practices. It is important to understand the specific requirements and challenges of these local decision-makers as well as the use of these data. The focus of

policymaking within a local setting, as a result of the introduction of the ICS, makes it timely to conduct this research, as data from the BaBi studies have the potential to inform these decisions.

Chapter 4: Identifying Research Priorities for Linked Data Research

4.1 Introduction

The identification and prioritisation of research topics is central to ensuring the BiB4All and BaBi cohorts have a notable impact on developing the evidence base related to child and maternal health. It is the first stage of the BaBi LHI model, detailed in Figure 10, and is part of the foundational phase of the FHI 360 Research Utilisation Framework.

Figure 10 BaBi Local Health Intelligence Model*

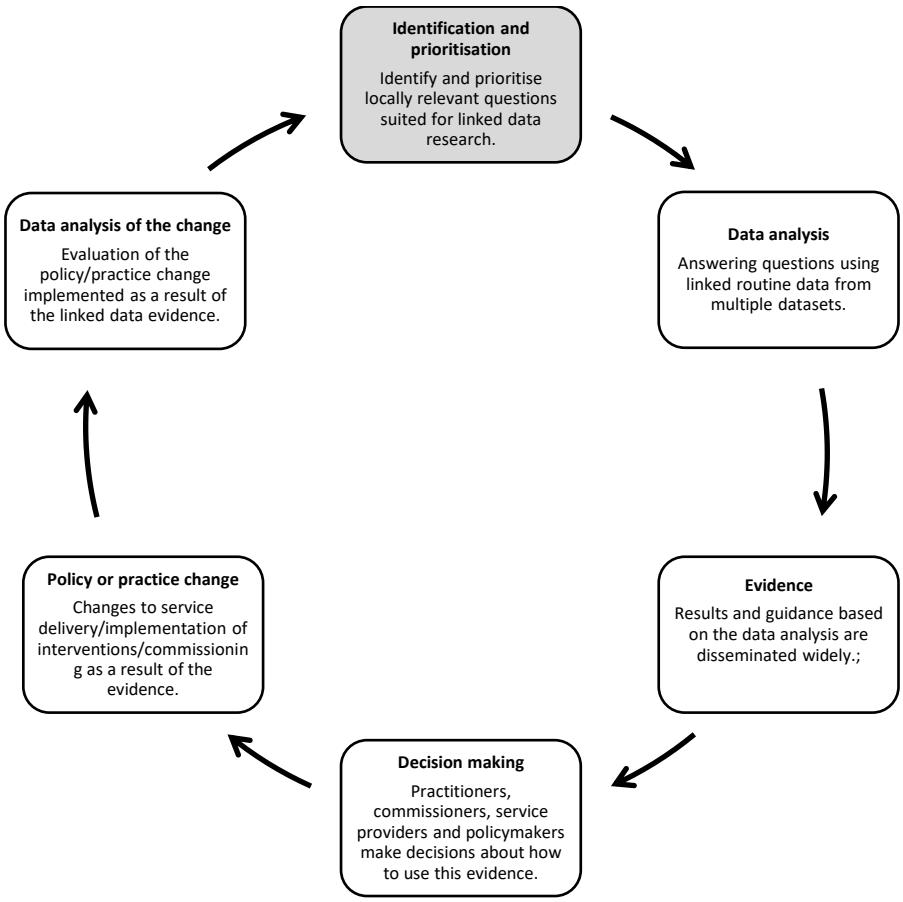


**Figure adapted from (Bryant and Bridges, 2021, Unpublished)*

This chapter describes the development of a method aiming to identify research priorities around child and maternal health, which are important locally, and can be addressed with the BaBi linked data.

Engagement of policymakers by researchers is often an essential part of the research translation process and is actively encouraged by most UK research funders (UKRI,

2023, Wellcome, 2023). Engaging policymakers in the setting of research priorities is one of the main activities as part of the foundational phase of the FHI 360 Research Utilisation Framework. Oliver *et al.,* (2019), suggest that co-producing research with relevant stakeholders has the potential to be more impactful as the research questions are designed with the implementation setting in mind. They also indicate that collaboration between researchers and research users can negate negative stereotypes about each other, allowing for more credible and relevant research. Moreover, Wright *et al.,* (2022) suggest that early engagement across policy and practice is critical for maximising the benefits of linked datasets.

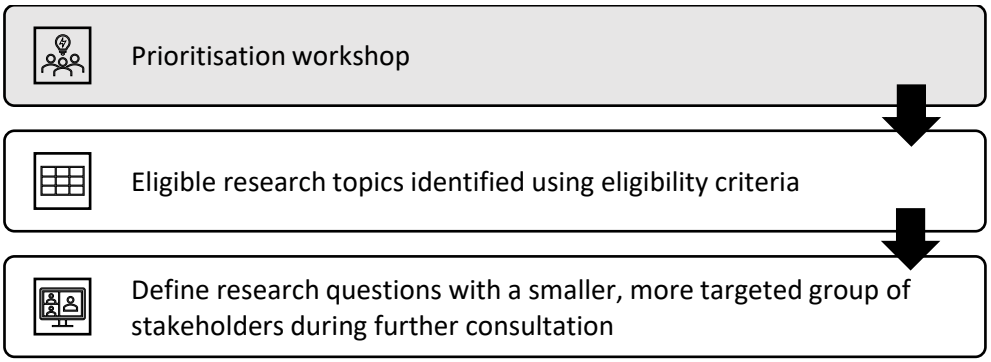
Considering the concepts outlined in Caplan's (1979) *'two-communities model'* of knowledge transfer, engaging policymakers in research can improve the mechanisms of communication and levels of trust between the research and policy communities. Alternatively, engaging stakeholders in the prioritisation of research can support the *'problem-solving model'* of knowledge transfer, whereby a policy problem is first recognised through engagement. This then stimulates research that aims to provide the evidence base for policy solutions. However, this linear model of the relationship between policymakers and researchers is often criticised and described as unrealistic by the research utilisation community (Smith, 2013). This thesis considers the relationship to be more complex than the linear model, and instead considers the interplay between research and policy. Hence, the aim of this chapter is to generate research priorities for linked data research, so that they can contribute to the policy evidence base and be used alongside other forms of evidence in local decision-making. This is consistent with the *'enlightenment model'*, proposed by Weiss (1979).

The mapping review, presented in Chapter 2, identified stakeholder engagement in the prioritisation of research as one of the strategies used by researchers in data linkage studies to promote the use of their research findings by decision-makers. However, the review revealed a gap in the knowledge regarding best practice for engaging stakeholders at this stage. This was consistent with what Oliver *et al.,* (2022) found when exploring research-policy engagement.

It was necessary to develop a method of engaging local stakeholders to identify research priorities for the BaBi studies, as existing priority setting processes were not appropriate. The method described in this chapter involves hosting an online prioritisation workshop, which brings together local stakeholders to discuss important areas of child and maternal health that can be explored with BaBi data. Chapter 5 of this thesis discusses how the research priorities identified during these prioritisation workshops, can be refined into actionable research questions that have the potential to be addressed with linked data

from the BiB4All study. This forms part of a priority setting process, which is summarised in Figure 11.

Figure 11 Overview of the priority setting process for BaBi research



This process can then inform the next stage of the BaBi LHI model, which is to use linked routine data to address an identified research priority.

This chapter begins by presenting the involvement and engagement literature that informed the development of the priority setting process. The prioritisation workshop method is then discussed and the outputs from an initial workshop are presented. Next, I discuss how the method was refined and applied by the BaBi pilot sites. This chapter concludes with a discussion of how the prioritisation workshop outputs inform the next stage of the BaBi LHI model, which is the focus of Chapter 5.

4.2 Public and Stakeholder Involvement

Historically, the health research agenda has been driven by researchers, meaning research often fails to address the needs of relevant stakeholders. The importance of public and stakeholder involvement in health care research is now widely recognised for helping to focus research on local priorities (INVOLVE, 2012).

Patient and Public Involvement (PPI) in research is referred to *as "research being carried out 'with' or 'by' members of the public rather than 'to', 'about' or 'for' them"* (INVOLVE, 2012 pg.6). The NIHR extends the remit of involvement beyond patients and the public to describe research as a joint venture between patients, the public, researchers, clinicians, and health professionals (Denegri, et al., 2015). Researchers increasingly want to engage with practitioners, policymakers, and members of the public, but face dilemmas regarding when and how to engage. A growing number of sources have

provided advice on engaging stakeholders in research, but there is no method of best practice identified.

The Goldacre review, which discussed in section 1.2, describes PPI as central to ensuring productive and ethical use of NHS data (Goldacre and Morely, 2022). Administrative Data Research (ADR) UK reviewed the literature on public attitudes towards sharing and linking routine data and they found that the public were supportive of the use of linked routine data for research if three conditions were met. One of these conditions related to public interest, where use of these data should demonstrate that it is in the public interest and has the potential to lead to tangible societal benefits (ADR UK, 2020). Therefore, engaging members of the public in linked data research, to set the research agenda, can ensure that these data are being used for research that is in the public's interest.

There are several levels of involvement, outlined by INVOLVE (2012), which are summarised in Box 9.

Box 9 The Levels of Involvement

Consultation: researchers ask the public/service users for their views and advice.

Collaboration/Co-production: researchers work collaboratively and in equal partnership with stakeholders to develop the research project.

User led/controlled: patients make the decisions about the research.

The term engagement is often used alongside the term involvement and describes how information and knowledge about research is shared (INVOLVE, 2012). Examples include dissemination of study findings to research participants, colleagues, and wider members of the public, and raising awareness of the research through the media and open days where people are invited to find out about research. In contrast to involvement and engagement, participation in research is where people take part in a research study and are part of the data collection process (INVOLVE, 2012).

Involvement can occur at different stages of the research process from identifying and prioritising research topics, through to disseminating the research. However, this thesis focuses on involvement at the start of the research process. Tong *et al.* (2019) explains that by involving stakeholders explicitly in research priority setting it can: (i) help ensure funding is directed towards research that meets critical evidence gaps, (ii) encourage

accountability and a shared responsibility for implementing the research, (iii) improve the relevance and legitimacy of research, and (iv) lead to better health outcomes.

Despite the potential benefits of public and stakeholder involvement, there are concerns about whether involvement and engagement activities always provide a return on investment as they can be costly in both time and money (Oliver and Cairney, 2013). It has also been highlighted that those who participate in involvement and engagement activities may not be representative of the wider population due to the large time commitment. Oliver and Cairney (2013) suggest that this could give rise to bias by favouring the views of those involved. Thus, involvement should be broken down into small, manageable components to reduce the burden on those involved (Denegri, et al., 2015). It may also hinder a researchers' ability to speak critically through fear of damaging the working relationship.

Oliver *et al*. (2017) outlined some potential challenges that are associated with coproducing research with the stakeholders. The main argument related to potential conflicts between involved parties, where heterogeneous stakeholders disagree on the direction of the research. In this case, confident stakeholders likely dominate the discussions, and quieter individuals may have less input. Therefore, when diverse stakeholders are involved in research and where there is likely to be a difference in opinion, careful facilitation is needed to ensure all voices are heard. In addition, Oliver *et al*. (2017) argued that the purpose of involvement and engagement is usually to achieve policy impact, which may never arise. Nonetheless, the benefits of public and stakeholder engagement in health research supersede the costs.

The Health Research Authority (HRA) sets out four principles for meaningful involvement of members of the public in health and social care research, which are to: (Health Research Authority, 2023):

1) Involve the right people,
2) Involve enough people,
3) Involve those people enough, and
4) Describe how the involvement helps the research.

Recent guidance published by the NIHR has been designed to promote and develop research practices that support effective patient and stakeholder involvement (INVOLVE, 2018). It sets out the UK National Standards for Public Involvement, as shown in Box 10, which can be used as a tool for evaluating the success of the involvement. However, it is acknowledged in the literature that it is difficult to trace the impact of involvement and engagement activities on research use by policymakers (see section 1.5.5). The design

of the engagement method developed in this thesis draws on the principles set out by the HRA and NIHR Standards for Public Involvement.

Box 10 UK National Standards for Public Involvement

Inclusive Opportunities
Offer public involvement opportunities that are accessible and that reach people and groups according to the research needs.

Work Together
Work together in a way that values all contributions, and that builds and sustains mutually respectful and productive relationships.

Support and learning
Offer and promote support and learning opportunities that build confidence and skills for public involvement in research.

Communications
Use plain language for well-timed and relevant communications, as part of involvement plans and activities.

Impact
Seek improvement by identifying and sharing the difference that public involvement makes to research.

Governance
Involve the public in research management, regulation, leadership, and decision

In response to the global Covid-19 pandemic, the NIHR agreed some new commitments for involvement and engagement. The commitments involved finding adaptive ways of communication such as digital or remote working, that were appropriate for contributors (Isaacs, 2020).

4.2.1 Priority setting in health care

Whilst there are a number of approaches for prioritising health research topics, there is no single best practice. In this section, I present several well documented approaches

for priority setting in health care research that utilise methods of involvement and engagement.

Consensus-based methods are frequently used to prioritise health research (Tong, et al., 2019). Nominal group technique is an example of a consensus-based method which engages experts in a highly structured meeting consisting of two rounds. The first round requires those involved to list their ideas on a topic without discussion. Each person involved then contributes an idea to the facilitator which is noted down. The ideas are grouped together, where appropriate, and discussed. Each person then privately ranks each of the ideas. After the results from round one have been tabulated and presented to the group, round two involves discussing the results followed by a re-ranking exercise (Jones and Hunter, 1995).

The James Lind Alliance (JLA) approach establishes Priority Setting Partnerships (PSPs), which bring together patients, carers, and clinicians to agree on the most important areas for research. The JLA process can take up to 18 months to complete. It involves forming a steering group to oversee the process, where they first gather research questions via online surveys from patients, carers, and health professionals. These research questions are then narrowed down by cross checking with current research and further surveys and individuals can vote on their most important question. Finally, a PSP workshop brings together patients, carers, and health professionals to jointly agree on the top 10 questions most important for research (NIHR, 2020b). The PSP workshop is a type of consensus method, where the main goal is to agree on priority research questions in specific areas of health. During the workshop, stakeholders use cards to rank the priorities from least to most important.

Alternatively, the Delphi method is often used for prioritisation of health research and involves administering a survey to a group of experts over several rounds. After each round, the results of the survey are reported back to the group, and the next round involves administering another survey. The process stops when there is a convergence in opinion or when a point of diminishing returns is reached. This process is not appropriate where personal contact and discussions amongst contributors are desirable as surveys are completed independently (Fink *et al.,* 1984).

The Child Health and Nutrition Research Initiative (CHNRI) have developed a systematic method of setting priorities for health research investments, to be used by international agencies, research funders, national governments, and policymakers. Each potential research investment is judged based on a set of criteria to assess the likelihood that each option can realistically contribute to reducing the burden of disease. It involves technical

experts independently placing scores next to each of the research options based on the specified criterion. This process allows for both idea generation and prioritisation (Rudan *et al.,* 2008). Similar to the JLA approach, this is a time-consuming process and is not appropriate where rapid priority setting is required. The CHNRI method provides a useful process for developing a criterion for which to prioritise research questions.

Other methods of collecting and selecting research priorities include interviews, focus groups, workshops, and surveys, where these may be conducted online or face-to-face (Tong et al., 2019).

Each of the approaches described in this section were considered for identifying and prioritising research as part of the BaBi LHI model however, no single method was appropriate.

Firstly, the priority setting method needed to be rapid to ensure the research remains relevant in a fast-changing policy environment. Both the JLA and CHNRI approaches are time-consuming, meaning that the priorities generated can quickly become outdated. It was also important that the approach allowed for group discussions. BaBi links data across health and other public services, and a group discussion would allow the stakeholders involved to share expertise regarding the data that are collected as part of their service. It would also enable researchers to gauge current understanding of linked data amongst stakeholders and members of the public, which could be used to inform future engagement plans. This is important as data linkage is complex and relatively new to most local areas. These discussions would be unobservable in methods such as JLA, where the research ideas are produced in advance of the workshop, or in the Delphi process where experts are consulted through survey methods. Moreover, the priority setting approach needed to be appropriate for busy stakeholders to attend and could be easily applied by teams setting up BaBi studies. The teams setting up BaBi studies as part of the pilot were mainly formed of clinical experts who had less experience in conducting research. Hence, the method needed to be suitable for these teams to use. I took inspiration from the existing priority setting methods to develop a pragmatic approach that considered the needs of the BaBi LHI model but was grounded in established processes.

The next section discusses the development of an inclusive method that seeks to involve members of the public and stakeholders in a rapid priority setting activity, that can be used in the first stage of the BaBi LHI model. Each stakeholder group within the early years space is likely to have different priorities, reflecting their different backgrounds.

Hence, it was important that the prioritisation process engages and accommodates these different perspectives.

4.3 Aim of this chapter

The aim of this chapter was to develop a method of engaging local stakeholders and members of the public, to identify research priorities that could be addressed using data from the BaBi studies. The intention was for the prioritised research topics to inform the next stage of the BaBi LHI model, which is addressed in Chapter 5 of this thesis.

In the process of identifying local research priorities for the BaBi studies, I also aimed to explore whether there any challenges which may explain why limited evidence of linked data being used in decision-making was identified in Chapter 2.

4.4 Developing the method

The method developed in this chapter was informed by the literature set out in section 4.2 and supported by specialist public involvement officers from the NIHR YHARC who helped ensure inclusivity. It was also developed in collaboration with the BiB4All and BaBi management teams as they are the end users of this method.

Former one-to-one engagement of senior clinicians by members of the BiB4All central management team revealed that health professionals needed to be supported to develop their ideas for linked data research. It was apparent that although they saw value in the cohort, they were unsure how they wanted these data to be used. Hence, prior to the start of my PhD, members of the BiB4All central management team were awarded funding from the Research Design Service (RDS) Yorkshire and Humber Public Involvement Fund to conduct PPI activities for the BiB4All project. Sally Bridges (SB) and Professor Maria Bryant (MB) invited me to collaborate and utilise these funds. SB and MB articulated their ideas for how the funding could be used and I further developed and operationalised these ideas through designing and hosting a prioritisation workshop. Following on from a successful first workshop, we discussed how developing BaBi sites could also benefit from hosting a prioritisation workshop.

Identifying local research priorities during the set-up phase of a BaBi site can help prioritise datasets that should be linked to address these questions. It can inform data access requests and data sharing agreements as it facilitates understanding around data that are important and urgent to access. Identifying research priorities once recruitment

for a local BaBi site has begun, can help direct efforts towards research topics that are important locally. It can also serve as an engagement method to bring together stakeholders around the project. Moreover, by using a consistent approach, research priorities from across local areas can be brought together to guide research carried out with the BaBi meta-cohort. Therefore, I supported the BaBi pilot sites to apply this method during the initial developmental stages of their BaBi studies. I supported BaBi teams in Doncaster, Wakefield, Leeds, and East London through the process of organising and hosting a workshop, which allowed the method to be developed further based on these experiences. The local BaBi teams were able to adapt the method to suit their needs and were responsible for hosting the workshop.

Following these workshops, the method was written up and was included as a section in the BaBi Toolkit (see section 3.2.2.3). The toolkit is written in lay language and accompanied by a training video that I produced. The training video aimed to be a more accessible way for BaBi sites to learn how to apply the method, and this video can be made available on request.

The prioritisation of research for child and maternal health is a huge task, given the broad scope. Techniques were also limited due to the ongoing Covid-19 pandemic. Hence, I took a pragmatic approach to maximise the use of resources and work effectively with stakeholders and members of the public, given the context. A two-hour online workshop method was developed, based on the advice of local stakeholders.

Hosting the workshops online was appropriate as it facilitated bringing together previously disconnected health and care professionals and members of the public, who shared the common goal of improving the health of their community. However, online methods are not without their challenges. Online meetings have the potential to exclude individuals who do not have access to technology or those who don't feel comfortable using it. This could lead to underrepresented groups in online involvement activities (Irani, 2018). The intended attendees of this workshop were likely technology literate due to their professional roles in health and care and many people (including public contributors) had adapted to using videoconferencing to communicate as a result of the Covid-19 pandemic.

Evidence has suggested that working for an extended period of time online can require more concentration, resulting in online fatigue (NCCPE, 2020). Therefore, it was important to offer breaks during the workshop and keep the meeting as short as possible. Recent studies showed success in holding 2.5 hr workshops and this influenced the timings of this workshop (Forbes *et al.,* 2022). Furthermore, technical difficulties and poor

internet connections can distract from the original purpose of the meeting. To mitigate technology issues, a user guide for the online platform and some tips for improving internet connection were circulated to attendees prior to the workshop. Extra time was also allotted for workshop tasks, in the event that those issues occurred. I also made my contact details available for workshop attendees, in case of any problems.

The videoconferencing platform Zoom (https://zoom.us) was chosen to host the online workshops. As the University of York provides students with a premium Zoom account, I was able to access the full range of features. BiB also had access to a premium account, which supported local BaBi sites to host workshops on this platform. Zoom is user-friendly and individuals do not need an account to join the meeting. It can support large meetings; place attendees into small groups; record the meeting without the use of third-party software; and allow the meeting host and co-hosts to share their computer screen with other people at the meeting. Zoom also has a waiting room feature that ensures only invited attendees are admitted into the meeting (Archibald, *et al*. 2019). The literature on the advantages of using Zoom for research (Archibald, *et al*., 2019; Irani, 2019; Lobe *et al.,* 2020) further strengthened the decision to carry out the workshop this way. New BaBi sites may choose to use a different videoconferencing platform, such as Microsoft Teams, which is supported by many NHS trusts (NHS, 2023b).

The need to involve a diverse range of stakeholders strongly influenced the design of the workshop as it needed to be inclusive and support contributions from all members. As data linkage is complex, researchers who have previously engaged stakeholders and members of the public in linked data research have recommended developing knowledge around linked data to allow for more effective engagement (Roblings *et al.,* 2021; Scottish Government, 2012a). Roblings *et al.,* (2021) suggested providing sufficient background information and holding more than one meeting to help build this understanding over time. Hosting multiple sessions with stakeholders and members of the public were considered, however, the Covid-19 pandemic had increased the pressure of those working in health and care services. Therefore, it was important that this method did not burden busy health professionals. Instead, the method I designed provided those who had the time and interest the opportunity to remain involved after the initial workshop.

In the initial development stages, a pilot workshop was conducted with a group of peers to test out the technology, the comprehensibility of the information provided to attendees prior to the workshop, and to ensure the planned tasks were achievable in the allocated time. Recommendations from this pilot were incorporated into the workshop design.

Feedback on this method was sought through Google forms following the initial BiB4All workshop and in the qualitative research detailed in Chapter 6. I asked those involved in planning and conducting the workshops for their views on what worked well and what could be improved.

4.5 Prioritisation workshop method

This section details the method I developed for organising and hosting a prioritisation workshop that has been applied by the BiB4All team and BaBi pilot sites. Local sites adapted the method to suit their needs, hence, any deviations from this method can be found in section 4.7.

4.5.1 Facilitators

Firstly, a team of people who are going to facilitate the workshop are identified. The workshop facilitators are an important part of ensuring the workshop is a success as they guide the group discussions. Facilitators are usually members of the BaBi team, but they can be anyone who is familiar with the BaBi project. Facilitators of the workshops detailed in this chapter included research midwives, PhD students, researchers, clinicians, and commissioners who work with the BaBi project. Table 17 summarises the facilitator roles and approximately how many people are needed to occupy those roles.

Table 17 Prioritisation workshop facilitator roles

Role	Explanation	Number of facilitators
Chair	Ensures the agenda is closely followed and can also be a small group facilitator.	One
Technical support	Is responsible for admitting people into the workshop from the Zoom waiting room, places attendees into their breakout groups for the workshop tasks and re-admits people to the session if they get disconnected.	One
Small group facilitators	Facilitate the small groups discussions throughout the workshop	Approximately between five and seven
Note takers (optional)	Notetakers may be needed to note down the discussions from the small groups if the team decides not to record those discussions using Zoom.	Approximately between five and seven

Facilitators are provided with training on the platforms used to host the workshops and a guidance document for reference during the session. The guidance document for facilitators can be found in Appendix B.

4.5.2 When to hold the workshop

Deciding when to host the workshop is usually based on facilitators' availability. The workshop should be held within the hours of the working day, avoiding school holiday and festive periods, and start and end times arranged to avoid school drop off and pick up. This ensures as many of those who wish to take part are supported to do so. Ideally, the date should be planned and advertised far enough in advance to align with rota requests for clinical stakeholders. The workshop lasts no more than two hours, with a break mid-way through the session to reduce the risk of online fatigue.

4.5.3 Intended attendees

Relevant stakeholders from a variety of backgrounds are invited to attend the workshop, where each have important perspectives towards topics on child and maternal health. At the time this workshop method was developed, the BiB4All cohort included mothers and children in their early years of life. Hence, relevant stakeholders included:

- Parents and members of the community who work with or represent parents.

- Early years health and care practitioners (which included midwives, health visitors, GPs, Neonatologists, paediatricians, obstetricians, gynaecologists and other specialists in child and maternal health).
- Local authority employees with a remit for health and wellbeing, Public Health specialists, early years education and support practitioners.
- Researchers.

As the BiB4All and BaBi cohorts develop, it might be appropriate to invite other stakeholders, e.g., educationalists.

With reference to HRA guidance for PPI, involvement is not about statistical representation of the population, but that it should be about capturing the breadth and depth of stakeholder views (Health Research Authority, 2023). This means capturing a broad enough range of views on the issues that are likely to be important to people who the research is intended for. Hence, the aim of the workshops is not to be representative of the whole population, but to represent a range of perspectives towards local evidence gaps, that allows the research priorities reflect a range of voices.

Inviting a range of stakeholders to a single workshop is important as people are often familiar with data that are routinely collected by their service but are not as familiar with what is collected by other services. For example, a midwife may know what data are collected in midwifery but may not know the specific data collected by health visitors. Therefore, by bringing multidisciplinary stakeholders together, it allows them to discuss these issues with a range of expertise, to prioritise answerable research topics.

However, there is a concern that professionals may dominate the discussion as they have the confidence to talk about the issues, which may overshadow any public contributors in the group. Hence, careful facilitation is needed to ensure each attendee has the opportunity to contribute.

Some local areas part of the BaBi Network preferred to engage members of the public and parents separately. They felt that a large workshop, with many health professionals and commissioners, may overwhelm public contributors and that they would gain more from one-to-one or small group engagement activities. However, for those that did invite parents and service users, it allowed their perspectives to be captured and discussed with local service providers. This facilitates a broader understanding of priorities locally.

Based on the learning gained from hosting the workshops in each local area, I would recommend inviting between 25 and 35 individuals to the workshop, with representatives

from a variety of backgrounds. The attendees will be split into smaller groups for the tasks. If more than 40 people are interested in attending the workshop, hosting more than one workshop should be considered. The challenges of hosting workshops with over 40 people are discussed later in this chapter. It is also important to get a balance of the number of stakeholders from each background. If there are many contributors from one professional background, this will likely steer the workshop outputs towards the knowledge, experiences, and preferences of that stakeholder group.

Potential contributors with relevant backgrounds can be identified by the local BaBi teams through personal contacts, existing public contributor groups, and utilising connections to local NHS trusts and local authority through their local steering groups.

Once identified, potential contributors should be invited to the session by email, at least one month in advance. Template email invitations can be found in Appendix B. The email invitation may include a link to an Eventbrite or similar page for people to register their interest in joining the workshop. Creating an Eventbrite or similar page allows the collection of information such as the person's professional background, the organisation where they work and their contact details, which is useful for organising the event. Alternatively, the invitation can ask people to email a named contact to express an interest in joining.

Once a potential contributor has expressed an interest in joining the workshop, they are sent a calendar invite, which includes the link to the Zoom meeting. In advance of the workshop, potential contributors are also provided with an information sheet which provides details on BiB4All and BaBi, outlines what workshop attendees can expect from the workshop and what the workshop outputs will be used for. The information sheet circulated as part of the BiB4All prioritisation workshop can be found in Appendix B. This gives attendees the opportunity to ask any questions they have before the session. Reading the supporting documents is not compulsory but is there to help those who want a bit more information before joining. This information is also explained at the beginning of the workshop. All materials are presented in accessible formats, in language comprehensible to a lay audience, which aligns with the UK National Standards for Public Involvement (INVOLVE, 2018). Contributors are also sent a reminder email two days before the workshop.

It is important to keep a log of who has expressed an interest in joining the workshop and their professional background, as this will help when organising the attendees into groups prior to the workshop. It will also ensure that only people who are invited are admitted to the meeting.

4.5.4 Hosting the workshop

Prior to the workshop, one of the facilitators organises the list of individuals who expressed an interest in joining the workshop into small groups for the workshop tasks. The first task involves placing attendees into homogenous groups of between four and six people from similar professional or experience backgrounds. For example, Group one may consist of midwives, Group two may be made up of parents and members of the public and Group three could consist of commissioners. Groups of between four and six are recommended to ensure all attendees have the chance to speak. The second task then places attendees into new heterogenous groups with people from different backgrounds. Each of these groups is supported by a small group facilitator. Organising the attendees into groups before the session allows for an efficient use of time during the workshop. These small group discussions are referred to in this chapter as *'breakout sessions'*.

It is worth noting that not everyone who signs up to attend the workshop will attend on the day and some people might turn up who were not expected. Hence, it is important to be flexible to accommodate the change in numbers. If fewer people join, groups can be collapsed so that there are fewer breakout sessions. I recommend organising those that attend into groups of at least four people to allow for a meaningful discussion. If someone who has not signed up clicks to join the meeting, facilitators can make use of the Zoom waiting room feature. Facilitators can send a message to the waiting room to confirm their identity.

Table 18 presents an example agenda for the prioritisation workshop. This is the agenda followed for the BiB4All workshop. It is recommended that facilitators join the workshop at least 15 minutes early, to allow any questions and technical issues to be resolved.

Table 18 Example workshop agenda

Time	Agenda Item
10:00	**Join the Zoom call**
10:05	**Welcome and Introductions** *Recording begins*
10:10	**Ground Rules**
10:15	**Ice Breaker**
10:20	**Background presentation on BiB4All**
10:30	**Opportunity to ask questions**
10:35	**Explanation of first group task**
10:40	**Breakout session one (Task one)** Discussion around areas of child and maternal health and that are important for research
11:00	**Whole group feedback and explanation of second group task**
11:15	**Short comfort break**
11:25	**Breakout session two (Task two)** Prioritisation of research topics identified in task 1, based on urgency and importance.
11:45	**Whole group feedback**
11:55	**Concluding remarks**
12:00	**End the session**

The workshop begins by welcoming attendees to the session and an introduction from the workshop Chair. Ground rules are then established, and it is important that all group members agree to these rules of mutual respect and confidentiality. Zoom etiquette is explained, including muting of microphones unless speaking, and attendees are reminded that the session is being recorded for note taking purposes.

Following this is a short ice breaker, which involves both attendees and facilitators turning off their cameras and turning them on to wave if they identify with a particular group i.e., *"turn your camera on and wave if you are a parent".* Care should be taken to ensure everyone is included in this task.

Then, a short background presentation is delivered explaining the background behind BiB4All and BaBi. The presentation includes a short introductory video, which is used in the recruitment of mothers into the BiB4All cohort, and a PowerPoint presentation with worked examples showing the types of questions that can be answered with a linked dataset. Emphasis is on how these questions can only be addressed by accessing data

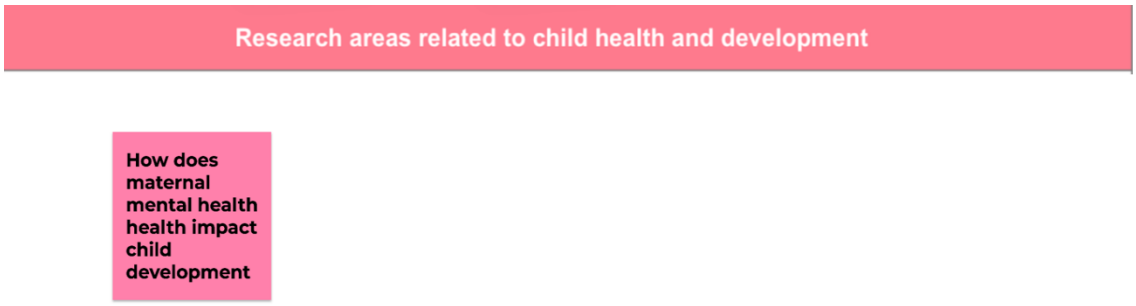
from multiple sources. This helps to develop public and stakeholder understanding of linked data so that they can effectively engage in the idea generation and priority setting process. This is important as previous research has revealed that people find it difficult to conceptualise the use of linked data (Roblings, *et al*. 2021). Presenters should ensure that the information is communicated in a way that everyone can understand. After this presentation, attendees can ask any questions they may have.

The remaining part of the workshop is split into two main sessions. The first part places attendees into the small groups that were arranged for task one prior to the session, utilising Zoom's breakout room feature. Once in the small groups, facilitators introduce themselves and begin the recording of the breakout room. Some BaBi sites opted to use notetakers or Dictaphones to record these discussions. In all instances, recordings and notes are stored securely, are only accessible by approved members of the research team, and are destroyed after the outputs have been written up.

Once the recording has started, facilitators invite attendees to introduce themselves. Attendees are then asked to discuss areas of child and maternal health they consider important for research and supported by facilitators to develop these ideas into research style questions with both an exposure and an outcome. It is recommended that facilitators encourage attendees to agree on between six and eight ideas as a group. Focusing on a small number of ideas ensures all ideas have the chance to be prioritised in the next part of the session, and allows the ideas to be developed. This task should last 20 minutes.

Facilitators make use of the online space, '*Google Jamboard',* to note down attendees' responses. An example is shown in Figure 12. Google Jamboard permits those with granted access to insert a virtual sticky note, mimicking how ideas would have been represented if the session had run face-to-face. Facilitators share their screen with attendees showing the Google Jamboard for session one. As the facilitator inputs the responses on the virtual sticky notes, facilitators can ask attendees to clarify their responses to ensure the research topics are clearly understood. A standardised format to record ideas is used, so that the ideas can be easily understood by all members of the team. This standard format is detailed in a guidance document for facilitators, alongside a list of prompts that can be used to stimulate discussion.

Figure 12 Google Jamboard Example



On the second slide of the Google Jamboard is a copy of the diagram showing the datasets that can be linked as part of BiB4All and BaBi, as shown in Figure 7. This can be accessed during this task, if needed, as members at the pilot workshop suggested that seeing the available data helped stimulate ideas. This also helps encourage participants to consider other datasets that might be useful for answering research questions.

Each facilitator is assigned their own Google Jamboard and allocated sticky note colour. This allows the research team to easily distinguish which stakeholder group generated each idea. This is useful for identifying similarities and differences in priorities across stakeholder groups. Copies of the Jamboards can be made available to the local BaBi team following the session, allowing all ideas to be considered. All the Jamboards are stored in a shared drive, making them easily accessible to the research team, and no personal information is stored on these boards. Google Jamboard is not commonly used within NHS trusts; therefore, training is essential for facilitators before each workshop. A selection of Google Jamboard templates are provided as part of the BaBi toolkit.

Attendees are then automatically returned to the main Zoom meeting. A volunteer of each group discusses their group's key ideas to the rest of the attendees. This gives all attendees and facilitators an idea of the key topics discussed, ready for the second half of the workshop. If there is a group of public contributors, I would recommend they have the opportunity to feedback first. This is followed by a short comfort break.

Whilst attendees are discussing their group ideas and during the short comfort break, one facilitator will be working behind the scenes. They will randomly allocate copies of

the virtual sticky notes generated in session one, to a new set of Google Jamboards that are distributed to facilitators for breakout session two. This ensures each group considers a wide range of ideas generated by people from different professional and experience backgrounds in the next task.

The second part of the session places attendees into new breakout groups, so that they are speaking to people from different groups. Each breakout group is allocated a facilitator and a Google Jamboard for breakout session two. Facilitators begin as they did in the first part, with a round of introductions and ensuring the breakout room is being recorded. Facilitators share their screen with the group to display the Google Jamboard for session two. Attendees are then asked to prioritise the allocated ideas, using the urgency and importance matrix shown in Figure 13. Importance is defined as something that has the potential for a large impact or is significant to improving child and maternal health outcomes. Urgency is defined as something that is time dependent or requires immediate action (Imperial College London, 2021). This method ensures consistency in the ranking of the ideas. It is emphasised to attendees that ideas ranked lower on the scale of importance and urgency will not be dismissed and that purpose the of the exercise is to decide which of the ideas should be addressed first. The facilitator can move the sticky notes to the quadrant suggested by the attendees and can edit the sticky note to add clarity to the research question, as required. There are 20 minutes allocated to this task.

Figure 13 Urgency and Importance Matrix*

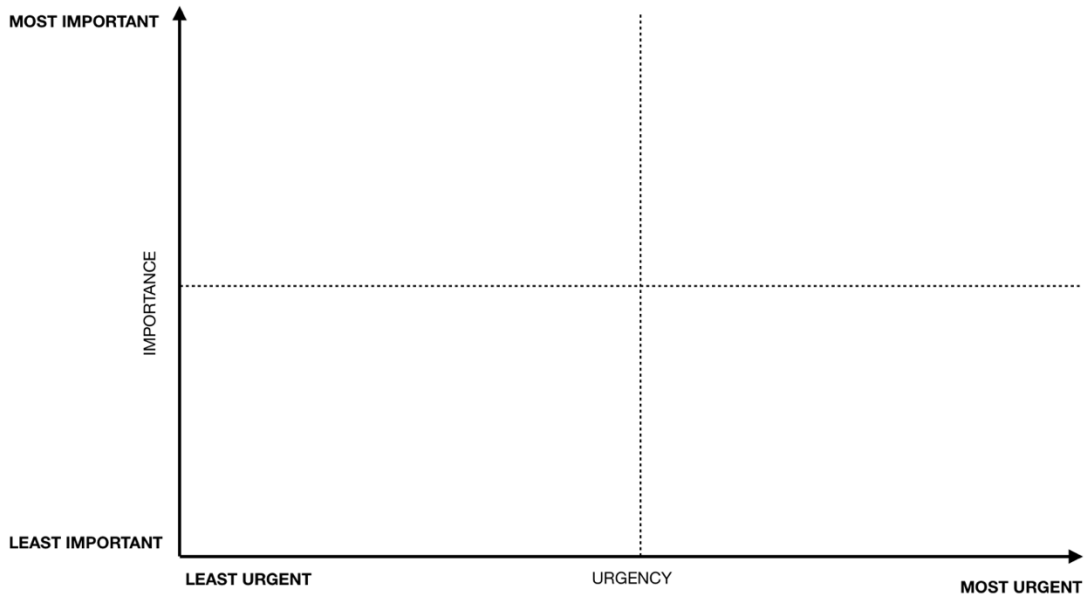


**Figure adapted from Imperial College London (2021).*

Following the second breakout session, attendees are returned to the main room and a volunteer from each group discusses their most important and urgent idea. The Chair then thanks everyone for attending the session and asks if anyone has any final remarks they would like to make. They also explain how the outputs from the session are going to be used and ask if attendees could fill out the feedback form for the session, where applicable.

4.5.5 Writing up the outputs

After hosting a prioritisation workshop, a summary report of the outputs is shared with the attendees and interested networks. These reports are used locally to guide research with the BaBi data. They could also be useful to those setting priorities more generally in the field of child and maternal health.

In the BaBi toolkit I have suggested that summary report describes:

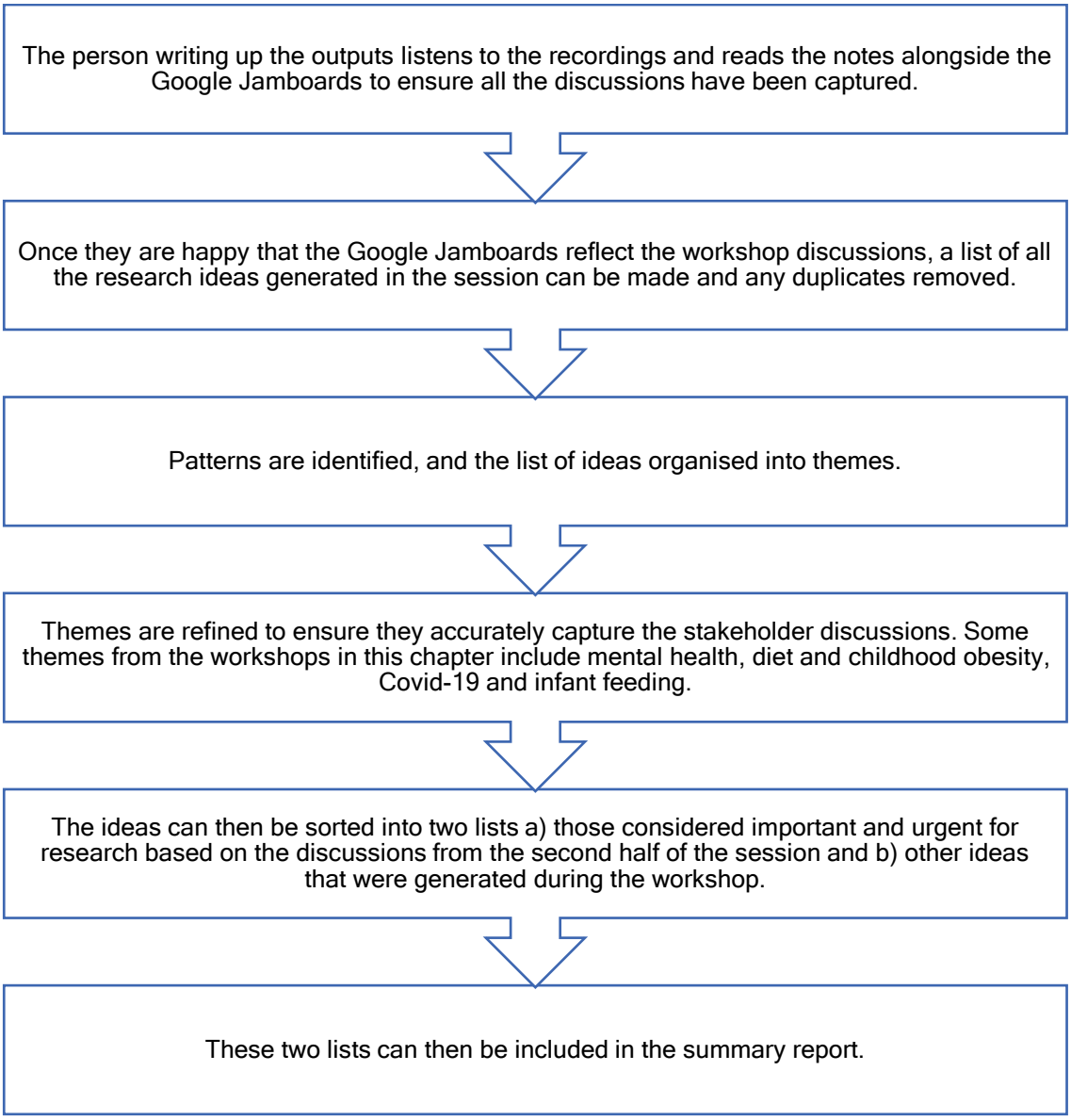
- The aim of the workshop.
- The stakeholder groups that attended.
- A brief outline of the session.
- A summary of the urgent and important ideas arranged under appropriate themes,
- a list of the other ideas raised during the workshops that were not indicated as urgent and important.
- Next steps and how the outputs will inform future research locally.

In reporting the outputs, care should be taken to ensure no contributor is identified in the outputs and it is clear how the discussions inform the research project, in line with public involvement and engagement standards (INVOLVE, 2018).

I recommend that a member of the BaBi research team or a researcher from a collaborating organisation writes up the workshop outputs. All the workshop recordings (including the Zoom recording of the main session and the breakout rooms and any Dictaphone recordings), notes, and Google Jamboards can be sent to that person.

Figure 14 represents the process I recommend for writing up the workshop outputs. It is based on Braun and Clarke's (2006) six-stage approach to thematic analysis, which is a method of capturing topics that emerge from a set of qualitative data.

Figure 14 Flow diagram demonstrating how workshop outputs were organised into themes



A consistent approach to writing up the workshop outputs allows the outputs from across different local areas to be brought together to identify research themes for the BaBi meta-cohort.

4.5.6 Ethical considerations

As this method aimed to seek opinions of local stakeholders and members of the public through involvement activities to inform the research process, ethical approval was not sought, although good ethical practices were followed throughout. The workshop method is not believed to raise any significant potential for physical and/or psychological harm to

the contributors or facilitators. In the event that an ethical issue does arise, appropriate procedures will be followed.

4.6 BiB4All workshop

The BiB4All prioritisation workshop took place on 18th Match 2021. The aim was to identify research priorities for the BiB4All cohort and to determine the feasibility of the workshop method. I was the lead on this workshop and was responsible for organising the workshop and writing up the outputs. I was the named contact for people should they have any questions or concerns and was mentored by a Public Involvement and Engagement lead within the NIHR YHARC to ensure those involved received adequate support during the process.

Attendees were members of the public and relevant local stakeholders from areas that are part of the BaBi Network (Bradford, Leeds, Doncaster, and Wakefield). This was the first application of the workshop method and learning from this experience informed subsequent workshops. At the time of hosting this workshop, BaBi East London were not part of the BaBi Network pilot and there was no representation from East London at the workshop. By hosting a workshop with attendees from across the BaBi sites, it allowed me to observe how different local areas were able to contribute and engage with the workshop to understand how it could be applied in different settings. Partners involved in developing BaBi cohorts also attended the workshop to learn how the workshop method can be used.

Involving stakeholders from multiple local areas also allowed a broader spectrum of opinions on the use of linked data and diversity in child and maternal health priorities across the region to be captured. The Covid-19 pandemic resulted in health service providers being even more time constrained, thus, the inclusion of the additional sites increased the population of potential contributors from each stakeholder group.

This initial workshop aimed to involve approximately thirty individuals with representation from a range of stakeholder groups. As this was the first workshop, the number of attendees was reviewed to inform future engagement. It was difficult to know in advance how many people would attend on the day, so more than thirty people were invited to ensure there would be adequate representation. I also ensured there would be enough facilitators if all those who expressed an interest attended. Strong links to the local community and stakeholder groups, established by the BiB team through their broader portfolio of work, were utilised for inviting relevant people from Bradford to the workshop.

Key members from each of the BaBi teams in Leeds, Wakefield, and Doncaster identified and invited potential contributors from their local areas.

The workshop was facilitated by five individuals with roles within the BiB4All central management team. They were familiar with the study and have backgrounds in research, data, or clinical practice, which made them suitable to facilitate the workshop.

In addition to this workshop, a smaller group session was offered to midwives who expressed an interest in the workshop but were unavailable on the chosen date. This extra session provided the opportunity for those unable to attend to contribute to the discussion and reflect on what was discussed in the workshop. Furthermore, a midwife from Bradford contributed her ideas via email and these were included in the final outputs.

Public contributors were reimbursed for their time according to the BiB payment policy for involvement in research. Public contributors were also supported to attend a Public Involvement in Health Research session to help build their confidence and skills in involvement and engagement. This is a free training programme developed with service users in the NIHR YHARC. It is a two-day programme aimed at service users, carers, and people who are new to public involvement in research (Richardson, et al. 2019). This supports inclusive opportunities as set out in the UK National Standards for Public Involvement.

4.6.1 BiB4All workshop outputs

The workshop was attended by thirty-four individuals from across Bradford, Leeds, Doncaster, Wakefield, and Sheffield, including representatives of those who are involved in the set-up of their local BaBi study. Table 19 shows the number of attendees interested in attending (column two) alongside the number of people that attended the workshop (column three) from each stakeholder group.

With reference to Table 19, seven individuals who initially expressed interest in the workshop did not subsequently attend. Midwives and public contributors expressed the most interest in attending the workshop, with the highest proportion of attendees being public contributors. At the workshop, there was approximately equal representation from each stakeholder group, although there were fewer clinicians than desired.

Table 19 Workshop attendees representing the stakeholder groups

Stakeholder group	Numbers registered to attend	Numbers attended
Midwives	9	5
Health visitors	4	4
Clinicians	4	3
Commissioners	5	5
Researchers	6	5
Public contributors	7	7
Other*	6	5
Total	41	34

*Other includes public health specialist, policy manager, project manager, business intelligence and a representative from a violence reduction unit.

Table 20 shows the number of people interested and the number of people who attended the workshop stratified by local area. The number of participants from each background was restricted to ensure a balanced viewpoint between the multiple stakeholders involved.

Table 20 Workshop attendees stratified by local area

Local area	Numbers registered to attend	Number attended
Bradford	20	18
Leeds	5	1
Wakefield	5	5
Doncaster	9	8
Sheffield	2	2
Total	41	34

Table 20 shows a high proportion of the attendees were from Bradford and that there was at least one representative from each of the BaBi sites. Two health visitor attendees, who worked as part of the 0-19 regional network for Yorkshire and Humber, recorded themselves as being from Sheffield. The 0-19 is a network of public health professionals working with the 0-19 age group (Institute of Health Visiting, 2021). Their attendance at the workshop was valuable as they were able to learn about data linkage, thus broadening the horizons of this study, as well as contributing insights from a health visitor perspective. In light of the workforce pressures, I was satisfied with the number of attendees who were able to engage in the session.

The workshop identified seventeen important and urgent research priorities which are presented in Box 11. Key themes included maternal and infant mental health, diet and childhood obesity, Covid-19, inequalities, infant feeding and labour and delivery. However, many of these priorities are not suitable for research using linked routine data, despite the facilitator guidance. For example, women's experiences of breastfeeding, including whether she stopped breastfeeding before she wanted to, are unlikely to be captured in their routine health record.

Box 11 Ideas indicated as most important and most urgent in the workshop

Mental health

- Explore the effects of women entering maternity services with pre-existing severe mental health on parent/infant relationships, bonding, prevention of separation.
- Investigate the impact of parental mental health on problems such as coping with sleeping and fussy eating.
- Examine the impact of parents with mental health difficulties not meeting the threshold for adult mental health services and the impact on family relationships and emotional wellbeing. Consider the impact of implementing a therapist that can provide links to crucial services?
- Explore the impact of early support (in the perinatal period and early years) on child and parental mental health. Discussion focused on adverse childhood experiences (ACEs) and parental experience with trauma and adversity as this may impact their parental ability and be emotionally available to the baby.
- The impact of early years and pregnancy on early years development, where it a comparison between Child Parent Psychotherapy (CPP) and usual care could be considered.
- The effect of mental health during the neonatal period on emotional attachment needs, breastfeeding, interaction, bonding and child emotional and physical development and school readiness.

Diet and childhood obesity

- The impact of proximity between take-aways and households on childhood obesity.

Covid-19

- The impact of access to limited support during the perinatal period and birth and how these impacts on maternal mental health and child development.
- The impact of social isolation in Covid-19, missed nursery time and play time on babies and infants forming meaningful relationships.
- The impact of Covid-19 on environmental allergies.
- The impact of Covid-19 on the likelihood of having a breech baby, linking in with rates of elective sections.

Inequalities and access

- Explore why Black and Minority Ethnic (BAME) communities are not accessing health care services and the barriers influencing this.
- Evaluate the effectiveness of online antenatal care and education compared with face- to-face and compare across the region.

Infant feeding

- Explore the impact of infant feeding on maternal mental health, meaningful mother/baby relationships (secure and insecure attachment) and long -term child health outcomes.
- The impact of reduced support for infant feeding, as a result of Covid-19, on weaning.
- Maternal experience of breastfeeding, including whether she stopped breastfeeding before she wanted to, access to support services, and how this impacted on the mother/baby relationship.

Labour and delivery

- Explore the factors influencing the likelihood of having a breech baby and the effects of a breech baby such as cost implications and the health of the mother and baby.

A summary of the workshop outputs was shared with attendees within three months of the session. This summary document clearly outlined how these outputs directly informed my PhD research and how the remaining research ideas formed a pipeline of projects to be conducted as part of BiB4All and the Early Life and Prevention (ELP) theme of the YHARC Collaboration. I also discussed the outputs from the workshop with other academics in the field of linked data, who were interested in applying this method.

As a researcher, it is important to continually reflect on how our own assumptions affect the outcomes of our activities. I began this process with pre-existing ideas about what the important areas of child and maternal health might be, which were informed by my prior reading and involvement with the team at BiB. Despite these preconceptions, when facilitating discussions with attendees, I was keen for attendees to express their views openly and without judgement or influence. The other workshop facilitators also had prior beliefs about important areas of child and maternal health for research. This may have influenced the way the discussions were directed in the breakout sessions. Facilitators endeavoured to remain neutral and allow the attendees direct the conversation towards

the areas they felt were most important. Moreover, the workshop was conducted during the Covid-19 pandemic which likely impacted the research priorities that were generated.

4.6.2 Feedback on the BiB4All prioritisation workshop

Feedback on this workshop was obtained via a Google form, which was circulated following the initial prioritisation workshop. The Google form also included an option to leave contact details, allowing attendees to be involved further in my PhD project if they would like to. Of the thirty-four people who attended, fourteen people filled out the feedback form, where thirteen people left contact details for future involvement. The feedback was mostly positive with individuals commenting that it was well organised. Some individuals felt more time was needed in the breakout rooms, however, others felt the time was adequate. Those who filled out the feedback form felt that the pre-session information was clear and timely and when asked if they enjoyed the workshop, all answered *'yes'.* This feedback informed the design of subsequent workshops detailed in section 4.7.

4.7 Application of the workshop method in other BaBi pilot sites

Following the success of the BiB4All workshop, I supported teams within the BaBi Network to host prioritisation workshops in their local areas. We worked together to develop, refine, and adapt the method as appropriate. Table 21 summarises the outputs from these workshops, which were written up by members of the local BaBi teams and details any deviations from the workshop method detailed in section 4.5. Each BaBi site decided how to present their workshop outputs, hence, there was variation in the information available to produce Table 21. For some workshops, there was only information related to the key themes available, whereas for other workshops there was detailed information related to the identified research questions. Each BaBi site has a list of their research questions, however, this was not available to include in this table.

As more of the BaBi teams successfully applied this method, there was greater interest from other teams in implementing and developing this approach. Each local area part of the BaBi Network pilot now has a list of research priorities, which will inform research projects within their local trusts and across the wider BaBi meta-cohort. A successful application of this method in these sites suggests that the method is transferable and can be adopted by other local areas joining the BaBi Network.

Table 21 A summary of the workshops in BaBi pilot sites

BaBi workshop	Attendees	Key ideas/themes	Deviations from the method in section 4.5
BaBi Doncaster - 10th June 2021 (Marvin-Dowell, 2021, Unpublished)	Attended by 18 delegates from Doncaster.	**Mental health** • Explore the environmental and social factors which impact upon child mental health and identify interventions to reduce prevalence of poor mental health among children and young people. • Investigate the role of attachment and other factors on neonatal mental health. • Investigate the longer-term outcomes of poor neonatal mental health. • Explore the issues around transition to parenting for first time parents. • Explore the relationship between birth trauma and maternal mental health. • Investigate the practices of health professionals (e.g., induction of labour, assisted delivery, caesarean section) in relation to women's experience of birth trauma. **School Readiness**	• Public contributors were not invited to this session and were engaged separately after the media launch of BaBi Doncaster in July 2022. • The first part of the workshop placed attendees into multi-disciplinary groups rather than single-disciplinary groups. • Facilitators were assigned note takers to type up the ideas on the Jamboards. • The first group task extended to 35 minutes.

BaBi workshop	Attendees	Key ideas/themes	Deviations from the method in section 4.5
		• In what ways can families be best supported to optimise social development and early language in children aged 0-4? **Childhood obesity** • To what extent is obesity transferred across generations of families? • Explore interventions for preventing the intergenerational impact of excess weight. **Services** • What are the barriers and facilitators for registration and attendance for dentistry? • What are family's experiences of the accessibility of family hubs and what impact have these had locally? **Breastfeeding** • What local factors impact upon initiation and duration of breastfeeding?	

BaBi workshop	Attendees	Key ideas/themes	Deviations from the method in section 4.5
BaBi Wakefield - 16th November 2021 (Hirst, 2022, Unpublished).	Sixty-four people attended the workshop, and they were predominantly from Wakefield as well as Leeds, Huddersfield, Bradford, Sheffield, and London. Attendees were from a variety of backgrounds including parents, specialist midwives and nurses, GPs, maternity and children's service commissioners, researchers, paediatric and neonatal specialist clinicians, community engagement officers, research delivery and governance specialists, public health knowledge and intelligence professionals and Yorkshire Sport Foundation representatives	• Maternal mental health • Mental health of children & young people • Infant feeding • Infant & child development • Raised BMI & obesity • Diet, nutrition & physical activity • Childhood illness & disease • Genetics & epigenetics • Gestational diabetes • Sleep • Deprivation & health outcomes • Language needs • Ethnicity & culture • Smoking • Drugs & alcohol • Housing & local spaces • Out-of-hospital birth • Parenting • Public health messaging & technology • Covid-19	

BaBi workshop	Attendees	Key ideas/themes	Deviations from the method in section 4.5
BaBi Leeds- 11th May 2022 (Hardicre, 2022, Unpublished)	The workshop was attended by thirty-eight delegates from across Leeds and Yorkshire. Delegates were from the following roles: parents, clinicians, midwives and nursing staff, health visitors, service commissioners and service design and delivery professionals, health improvement specialists, public health practitioners, local authority representatives, third sector organisation representatives, and researchers and academics.	• Mental health • Obesity • Diabetes • Personal and social risk factors • Services and resources: availability, access, and use • Substance (mis)use • Physical (in)activity • Covid-19 • Oral health • Foetal development and birth	

BaBi East London – July 2022
(Hindes, 2022, Unpublished)
	The workshop was attended by members of the public, service users, researchers and various professionals from health and social care (including midwives, health visitors, nursing staff, doctors, local authority representatives, academics, public health practitioners, service commissioners, and service design and delivery professionals).	• Service provision: access and availability
		• Health equity
		• Social and environmental determinants of health
		  ○ Food accessibility
		  ○ Housing and structural accessibility
		• Effects of Covid-19 and lockdown
		• Maternal health and healthy childhood Development
		• Long-term Child Health Outcomes	

The benefit of having single disciplinary groups for the first part and multidisciplinary groups for the second part of the workshop, is that each group in the second part has at least one representative from each of the groups from the first part. For example, each group in the second part of the session has a midwife, a public contributor, a commissioner, etc, in their group. Hence, when the ideas are randomised for the second part, there is a contributor in each group that can give context behind the ideas generated by their group in the first part. The BaBi Doncaster team decided that the groups for the first task would include a mixture of professional backgrounds, to explore whether this engaged more stakeholders, or promoted a broader discussion around data from different datasets. Reflecting on their decision, the BaBi Doncaster team decided that subsequent workshops would revert to having groups with similar backgrounds.

To reduce the burden on the facilitator in the breakout sessions, each facilitator in the BaBi Doncaster workshop was assigned a notetaker to type up the ideas on the Google Jamboards. This was to allow the facilitator to focus on supporting attendees to generate the ideas. I occupied a notetaker role during this workshop and found it challenging as I was unable to ask for clarification on the ideas, as the facilitator may have already moved onto another topic. Therefore, there were some instances where I needed to break the flow of the discussion to ask contributors to elaborate. Hence, I would recommend that the small group facilitator inputs the ideas onto the Google Jamboard.

Moreover, the BaBi Doncaster team extended the first group task from 20 minutes to 35 minutes, which reflected the feedback from the initial workshop. This resulted in lots of ideas being generated in this session. This made it difficult to ensure all ideas had the opportunity to be prioritised. Hence, in workshops where many ideas are generated, I would recommend that facilitators ask the group to identify their key six to eight ideas from this list. I would also recommend keeping the time for the first task at 20 minutes to ensure there is adequate time in the second task to prioritise the ideas generated.

The large number of attendees at the BaBi Wakefield workshop made the session difficult to manage. This resulted in the second part of the workshop being cut short and ideas were not able to be prioritised based on urgency and importance. As a result, the method detailed in section 4.5 recommends involving between 30 and 40 contributors to ensure there is enough time to complete both tasks. This also ensures all contributors have the opportunity to share their views during the session. This also explains why the workshop report produced by the BaBi Wakefield team presented the themes discussed in the workshop, rather than a list of important and urgent research questions.

The themes identified in the workshops detailed in Table 21 are comparable with the most important and urgent themes identified in the BiB4All workshop. Key themes identified across the workshops included: mental health, obesity, access to services, infant feeding, and Covid-19. This provides the opportunity to address some of these topics on a wider scale, utilising all the available data from the BiB4All and BaBi studies. Similar to the BiB4All workshop outputs, many of the research ideas identified in the BaBi Doncaster workshop were not suitable for research using linked data. For example, women's experiences of birth trauma are unlikely to be captured in their routine health record.

4.8 Workshop method feedback

During qualitative interviews with local decision-makers in areas setting up BaBi research, described in Chapter 6, I gathered feedback on the workshop method. Two sub-themes were identified that related to the engagement workshop: workshop format and workshop outputs. In this section, I summarise the key ideas that helped to improve the workshop method and develop the package of resources included in the BaBi toolkit. Overall, the feedback received regarding the workshop format and outputs was positive.

4.8.1 Workshop format

Many of the participants communicated a preference for face-to-face workshops.

> *"I do think there's nothing like face to face when you're doing something like that, though, because people then get to know the people as well"* (BaBi decision-maker).

Therefore, hosting face-to-face workshops could be considered. Although, participants did describe some benefits of communicating online such as attendees feeling more confident and that more people are able to join the session.

"*The good points being that more people can join, you know people communicate in different ways, don't they? And they may feel more able to communicate online via text than standing up in a large room, for example."* (Clinical decision-maker).

I would advise that local BaBi sites engage their local stakeholders in a way that best meets their needs.

Participants also discussed whether pre-workshop information should be provided to workshop attendees. It was suggested that pre-workshop information could allow attendees time to think about the topics of the workshop before attending. This was described by participants involved in hosting a workshop, as well as those who did not attend a workshop but have the potential to attend future workshops.

> *"I think it would be helpful to know ahead of the workshop, that that's what you're going to be asked… so that you could start thinking, ok, so this is about routinely collected data and questions that we could answer by looking at this routinely collected data. So, what do I think? What do I think, first of all, are the routinely collected data? And then to have a little think about what the potential questions could be?* (Clinical decision-maker).

However, it was also suggested that giving too much information may deter people from attending. Thus, there is a trade-off between giving enough information that people are able to engage with the research and giving too much that it discourages people from attending.

> *"If you send people in advance, that could also be off putting, because they think, Oh, I'm gonna go and talk about this and there'll be lots of other people who know more about it than me, and therefore I don't really want to, to turn up."* (Commissioner).

Pre-workshop information was provided to attendees of the engagement workshops and the feedback suggests that a balance is needed in how much information to provide.

Building up engagement with stakeholders over time was also discussed. For example, hosting small group meetings to help those who are new to research.

> *"But some people didn't know what they were coming to. That's the feedback I got from, let's say, some of the community midwives, they've never done anything like that before. So, for people that make policy for, you know, that's fine. They used to that. … But I know some of the practical clinicians are sort of like what am I meant to be saying? What are my ideas so I do think a big group session works in the end, but it might be just build it up, like little steps, little steps. Get a few engaged, give them a bit of a this is the overview. Could you just go away and think about it and then we'll bring you to a workshop"* (Clinician).

This is consistent with the recommendations of Roblings *et al.,* (2021) for engaging members of the public in linked data research. This suggests that continuous engagement of the same stakeholders is likely to be beneficial as they build up more knowledge of linked data research with each interaction. Increased engagement with the BaBi project can help give people the skills to contribute.

Based on these findings, it might be appropriate to provide more information about the workshop session, explaining exactly what they will be asked and how they can prepare for the session. A layered approach could be taken, whereby potential attendees are provided with brief information, allowing them to decide whether they would like to attend the session, and then more detailed information on linked data and the BaBi project can be made accessible, should attendees want to access this. This allows potential attendees to control the amount of information they access (Health Research Authority, 2019). I would also ensure that attendees have the opportunity to ask questions prior to the session as this can help reassure attendees and build their confidence in contributing during the workshop. It is important to emphasise that we are interested in a variety of opinions, to try and encourage attendance from those who might feel less confident in contributing.

People were positive about the workshop format. Those involved in hosting the workshops liked the use of Google Jamboard as it was easy for them to use. Therefore, I would recommend this technology continues to be used as part of the workshop format.

> *"I think the workshop was brilliant. I really enjoyed doing them and people said they were really good, Jamboards were excellent, because they're easy to use. And I think people could sort of see what they needed to do."* (BaBi decision-maker).

Finally, participants spoke about the benefits of being placed into multidisciplinary groups for the task.

> *"I think you'll be able to bounce off each other in terms of ideas of how you might use that data when you do similar jobs but might do them in different places. So, I think that will be useful, but I think if you just have … that group, you won't know what else is available and what else gets collected. So, certainly around some of the health inequalities data, for example. I wouldn't necessarily know what gets collected by the council or by another sector, but it could be really useful information."* (Service planning decision-maker).

As discussed, the BaBi Doncaster team decided to have multidisciplinary groups for the first task to encourage attendees to share their knowledge about their own datasets. This created challenges in organising the groups for the second task. Thus, the current workshop format is recommended as there is still the opportunity for people to discuss the ideas with people from different professional backgrounds.

Future workshops could consider three parts to the workshop. The first part could place people into groups with people from similar backgrounds to start the idea generation process. The second stage could place people into new groups with people from different backgrounds to develop these ideas, allowing people to share expertise on how you could make the most of data from other sources to investigate the research topics. The final part could then prioritise these ideas in multidisciplinary groups. This new format could increase the likelihood of the research topics making use of linked data, rather than data from one source. This could also allow attendees to build up the confidence to contribute by starting the session in a group with people from similar backgrounds, although this would require a longer workshop, which may not be feasible for busy health professionals. Hence, there is the potential for this workshop method to evolve, however, such variations would need to be tested.

4.8.2 Workshop outputs

When asked about the engagement workshops, many participants discussed the usefulness of the workshop outputs. A BaBi decision-maker said the workshop outputs are *"useful for different people for different reasons… our academic partners found them helpful from an academic perspective about what they might do and how they might do it, and what outputs they might use from it. .... And commissioners, I think, got a different view on it as to, 'oh not thought about that, how might we use this to think about what we do with services in the future?' And then I think providers, depending on the provider, I think they will some of us thought about, right, well, how do I get as many women as possible to sign up to this with families to see this build and build and build because it can only be a positive for our longer-term investment in health, delivery of health care, and the services we provide."* (BaBi decision-maker).

The view that the workshop outputs were useful for commissioners in thinking about their priorities was described by multiple participants.

> *"I found the whole research prioritisation process really useful… I really liked the fact that there are some outputs that allow us to have a real sense of consensus*

about what's really important….it has made me think that we … don't need to stop here with trying to work out what ongoing priorities are and that then it may well be that in some of those priority areas, we needed an additional deep dive" (Commissioner).

This supports the *'knowledge-driven model'* of research utilisation (see section 1.5.2), as the knowledge from this workshop provides the necessary pressure for policy to develop in line with this knowledge. In this case, it is to develop policies addressing some of the concerns around child and maternal health identified in the workshop. In relation to the *'enlightenment model'*, the workshops have influenced commissioners thinking towards priority areas for child and maternal health as well as linked data research. This influence, and their active involvement in the research process, makes it more likely that the research will be utilised by decision-makers as they are now considering these priority areas, and the research that is conducted as a result will contribute to the evidence base on this topic.

The workshop also has benefits for local BaBi steering groups in terms of broader engagement. One participant talked about how they expanded their steering group, as a result of the workshop, making this more representative of local services.

"And from the workshops, people then joined our partnership group. That's how we got [a clinical representative], [they] came to a workshop… And then from that, I emailed [them] to say, thanks for coming to the workshop…I'm really approaching you now to see whether you might be interested in being a part of our partnership group (BaBi decision-maker).

It was also suggested that the workshops allowed people to connect, which supported new partnerships to develop. This has resulted in additional research with BaBi. Therefore, hosting a prioritisation will be recommended for new BaBi sites.

"I think another hidden consequence positively is that people connected, so people didn't know, well they did know people, and they sort of knew names and so they either were like 'oh, I didn't know that was you' so that's great, or they'd never met, but actually from it, have done subsequent work in partnership... So, I think there are really positive unintended consequences of the workshop" (BaBi decision-maker).

4.9 Impact

To categorise the types of impact from the workshops, I selected some of the domains associated with the *'making Visible the ImpaCT Of Research'* (VICTOR) tool, developed by the Yorkshire and Humber Clinical Research Network and YHARC (Taylor, 2023). VICTOR is a tool to collect and make visible the impact of research activity within NHS organisations and is being used to record the impact of the BaBi project. Table 22 summaries the VICTOR domains. Table 23 presents a summary of the impact from each of the workshops using these domains.

Table 22 A summary of the VICTOR domains*

Knowledge Generation and Knowledge Exchange	Health benefits, Safety, Quality for Research Participants and Carers	Economic Impacts	Patient Service and Workforce Impacts	Research Profile and Capacity Building	Influences
Dissemination	Health benefits	Cost effectiveness	Service change	Research culture	Cohesion
Knowledge sharing	Patients experience	Cost saving	Skills	Research awareness	Reputation
New Knowledge	Patient safety	Commercialisation	Workforce	Research capacity	Recruitment and retention of staff
Actionable outputs	Equity	Commercial income	Collective Action	Networks and collaboration	Patient and public involvement
	Social capital		Products, equipment, technology	Engagement	
			Guidelines		

* *Table adapted from Taylor (2023).*

Table 23 Impact of hosting the workshops

Type of Impact	Description
Research capacity	The BiB4All workshop was attended by partners in the BaBi Network. The success of this workshop motivated teams in Doncaster, Leeds, and Wakefield to host their own workshops with support from the BiB4All team.
Research capacity	In November 2022, the outputs from the BiB4All workshop were shared with the Connected Bradford team as they had analysts transforming routine data and wanted questions to support the testing of the data.
Relationships	Through supporting the BaBi teams in Leeds, Doncaster Wakefield, and East London to host their workshops, I got positive feedback on the method and developed a network of contacts which I was able to utilise for my qualitative research, detailed in Chapter 6.

Type of Impact	Description
Relationships and awareness	As a result of the BaBi Doncaster workshop they were able to expand their local BaBi steering group to include a more diverse range of stakeholders. The workshop increased awareness, understanding and interest in the BaBi Doncaster project amongst a broader range of stakeholders. This broader reach and increased interest in the project enabled them to invite more members into their steering group.
Awareness	As a result of the BaBi Leeds Workshop, colleagues approached the coordinating centre, BiB4All, to be involved in shaping research.
Awareness	One attendee at the BiB4All workshop asked for permission to share the outputs with regional and national priority setting groups for 0-5s. They shared the outputs with senior members of the Institute for Health Visiting, Health Education England Yorkshire and Humber, Fuse (The Centre for Translational Research in Public Health), College of Nurses, as well as researchers at Newcastle University and a lecturer in Public Health at the University of Sheffield.
Knowledge transfer	Commissioners in Doncaster and Wakefield expressed during the qualitative research in Chapter 6 that attending the workshop encouraged them to think more about how the data can be used to support their decision-making.
Skills	Facilitators and attendees developed an understanding of how linked data can be used to address issues relating to child and maternal health locally.
Dissemination	Each workshop shared their outputs with those who contributed to the workshops and to the wider BaBi Network. Outputs from the BiB4All workshop were shared with Connected Bradford and the ELP theme of the YHARC.
Dissemination	In April 2023, the BiB4All outputs were shared with clinical colleagues by Sally Bridges.
Actionable outputs	The BaBi Doncaster team used the outputs from their workshop to inform their minimum dataset that will be available for research. They reviewed the outcomes that were most important and urgent and explored the data available on these core outcomes, much like what I describe in Chapter 5.

To capture these impacts, I regularly monitored the impact repository stored centrally on the BaBi Google Drive, where each BaBi pilot site inputs any impacts resulting from their activities. I asked those hosting the workshops to send email updates on how the workshops shaped their priority setting process as well as the unintended consequences of hosting the workshops. These unintended impacts included increased interest in being involved in the project or wider distribution of the workshop outputs.

In addition, as part of the Local Data Accelerator funding (see section 3.2.2.3), Katie Marvin-Dowell, employed by Doncaster Council, collated the reports from the research prioritisation workshops in these local areas and analysed these thematically using NVivo. The aim was to provide a topic focused picture of the areas which were of highest priority. This gave an indication of the priority areas for these BaBi sites.

Finally, on the 25th October 2022, I presented this method to the independent BaBi Network National Steering group which strategically lead the network.

4.10 Discussion

This chapter aimed to develop a method of engaging local stakeholders to identify research priorities for BaBi sites. The identified research priorities would also inform the research question addressed in Chapter 5 of this thesis, as part of the next stage of the BaBi LHI model. The developed method has been successfully applied by the BaBi pilot sites and has generated a list of important research areas for each of these sites. This demonstrates that it is possible to engage local stakeholders in the prioritisation of research as part of the BaBi LHI model.

However, many of the ideas generated during the workshops are not suitable for research using linked routine data and are better suited to qualitative research or research using bespoke survey data. Whilst these priorities can still be addressed with BaBi participants through the *'consent to contact'* mechanism, they will not be addressed as part of this thesis. This could suggest a need to build capacity and understanding about the potential of linked routine data amongst stakeholders, allowing them to better identify opportunities for these data to be used. This was reflected by the discussion in the BaBi Doncaster workshop where attendees discussed how they would like to understand more about what data are available in each dataset. Based on this, I would advise that local BaBi sites consider their own stakeholder groups' skillsets and knowledge about routine data research, so that they can adapt the prioritisation workshop method accordingly. For example, if there is very little knowledge and experience of linked routine data amongst local stakeholders, it may be appropriate to provide more information to workshop attendees prior and during the workshop about these data. Increased knowledge of data available as part of BaBi might lead to more research topics being generated that are suitable for linked data research or research using BaBi data.

The limited number of research priorities that were suitable for linked data research could also imply a need to build the skills of workshop facilitators, to allow them to identify ideas generated during the session that cannot be answered with routine data. This could allow facilitators to better guide attendees during the workshop to focus on research topics that are able to be addressed using linked routine data.

Moreover, this could suggest that routine data are not capturing information on key outcomes and exposures that are important locally. There could be a gap in routinely collected data, which limits the usefulness of these data to address local research priorities and be used as a local health intelligence tool for child and maternal health.

This provides an opportunity for local BaBi sites to consider whether there is additional information that could be collected routinely, to address these local research priorities.

However, I acknowledge that if facilitators limit the focus of the discussion to only research topics that can be addressed with the available data now, this could be a missed opportunity to identify gaps in routine data that could be improved by working with local services. Based on this, I would recommend developing an understanding of the data that are currently available in the BaBi datasets with local stakeholders, whilst highlighting that it is possible to work with local services and data providers to improve the collection of routine data for research.

Therefore, there are many potential uses of the workshop outputs, which include:

- Using the available data to address important and urgent research topics.
- Collecting additional data to explore questions that cannot be addressed using linked routine data. This could involve contacting BaBi participants to see if they would like to take part in research that collects these data, or BaBi teams could work with local services to improve the collection of routine data for research.

The research priorities identified during these workshops are specific to these consultation exercises and do not necessarily reflect the range of views in these local populations as a whole. The ideas represent the views of those who engaged in the workshops and reflect the context in which they collected. Thus, repeating this exercise with a different group of stakeholders and facilitators, at a later time point, may yield different research priorities. As attending a workshop was voluntary, it is likely that those who attended were interested in the project, meaning the priorities reflect the views of those interested people. In addition, the priorities likely reflect the stakeholders who are more confident in sharing their thoughts and speak more during the workshops.

Oliver *et al.,* (2022) discuss how engagement initiatives often compete with one another for decision-makers time. This makes it important to ensure non-tokenistic involvement, where attendees feel their ideas are being valued and considered. The priorities identified in these workshops were distributed widely to those who have an interest in child and maternal health. This means that other researchers can also benefit from this engagement activity, rather than duplicating through similar processes. As BiB utilises linked routine data for a number of projects, these outputs have been useful in informing research using routine data other than BiB4All, such as Connected Bradford. This prioritisation workshop involves a small time commitment, meaning it is unlikely to prevent attendees' engagement in other research projects.

There was a lot of enthusiasm for the BaBi project and there was lots of interest from people in attending the workshops. However, this enthusiasm could be because the BaBi project is new and exciting, and this may diminish over time. This could make it more difficult to host workshops in the future. To mitigate this, it is important to maintain momentum and keep people engaged throughout the project at all stages. This can allow rapport to be developed between the research team and public and stakeholder contributors. Involving the same contributors over time can help develop knowledge of linked data, better enabling them to engage with the research. Thus, to further strengthen this method, those organising the workshop could ask people to commit to being contacted after the session. This not only ensures continuous engagement of stakeholders in this research but allows researchers to contact attendees for clarification on the research priorities after the session, if needed.

4.10.1 Strengths and limitations

A key strength of this workshop design is that it supported a range of individuals, representing multiple voices and experiences in the maternity and early years space, to contribute to the research priority setting-process. Hence, the workshop priorities and subsequent research are relevant to a range of local services. In addition, the method is underpinned by the literature on priority setting for health research and the principles of meaningful public and stakeholder involvement and engagement. The method uses accessible concepts, languages, and presentation formats throughout and considers the skills and training needs of facilitators and attendees.

Compared to other documented priority setting processes, the method I designed required a significantly smaller time commitment from stakeholders, making it more accessible. The online platform also facilitated a wide range of people to attend a single meeting, despite the Covid-19 restrictions and the busy schedules of health professionals. It allowed parents to join without the need to leave their homes, which is potentially easier from a childcare perspective, and provided a relaxed environment. However, online methods are not without their challenges. As individuals attended from their home environment, there were more chances for attendees and facilitators to become distracted, such as the doorbell ringing or children/pets in the background. This resulted in some contributors not attending the whole session. The workshop method was designed with flexibility in mind so that attendees could easily join without having to commit to being involved for the full session, allowing more people to contribute.

A potential limitation of the workshops that were discussed in this chapter is that not all research questions suggested in the first part of the workshop had the opportunity to be prioritised in the second part. This is because there were lots of ideas generated in the first part. This meant that potentially important and urgent areas of research might have been missed. Therefore, the workshop design was amended to encourage facilitators to gain clarity on a smaller number of research ideas to ensure all ideas have an equal opportunity to be discussed in the second half of the session. Clarity of the research ideas is also important as many of the ideas from this initial workshop were broad, and by focusing on what is most important about that research idea, this makes the workshop outputs more useful for informing future research.

Reflecting on how the workshop outputs have been used, it would be beneficial to present the outputs of the workshops in standardised formats, as this allows themes to be compared across the workshops. The BiB4All outputs and those from the first BaBi Doncaster Workshop follow the format outlined in section 4.5.5. However, as the other workshops were written up in a slightly different format, this created challenges in comparing the outputs in the process developed by Katie Marvin-Dowell.

4.11 Chapter summary

The method detailed in this chapter forms part of a priority setting process, to identify and prioritise locally relevant research topics that can be addressed with data from the BaBi studies. Engaging stakeholders early in this research process, increases the likelihood that the outputs from BaBi research will contribute to policymaking. This is also a key part of the foundational phase of the FHI 360 Research Utilisation Framework (Kim *et al.,* 2018).

The prioritisation workshop method was underpinned by principles of meaningful public and stakeholder involvement and engagement, advice about the use of online platforms for research, and drew on established processes for priority setting for health research. I also considered the advice of Roblings, *et al*. (2021), who have previously involved members of the public in linked data research, in the design of this method.

Prioritisation workshops have been successfully hosted in five local areas within the BaBi Network, demonstrating that this method can be applied as part of a place-based approach. Applying this method in multiple BaBi sites, and gathering feedback from facilitators and attendees, also provided opportunities to develop the method further. To

support the implementation of this method, I wrote a section of the BaBi toolkit, which explained how to set up and run a pragmatic prioritisation workshop.

The prioritisation workshops detailed in this chapter were attended by a variety of stakeholders, representing many of the voices within the area of child and maternal health. The workshops highlighted a need to develop an understanding of linked routine data and its potential uses, amongst stakeholders and facilitators, to allow the workshop outputs to inform research using these data. Moreover, the workshops highlighted that linked routine data may not be capturing information on outcomes that are important locally, which influences how useful these data as local health intelligence for child and maternal health.

All the research ideas discussed in the workshops have been reported to workshop attendees and shared across relevant networks. The themes identified in the workshops have created a pipeline of potential projects for each local area, where common themes across the workshops can be addressed as part of the BaBi meta-cohort. The resulting list of priorities can also be a valuable resource for health researchers, commissioners, and other stakeholders investing in child and maternal health research across the Yorkshire and Humber region. The method I designed can be adapted by other research teams looking to identify and prioritise locally relevant research questions, in a timely manner.

The workshop outputs directly informed the research in Chapter 5, which aims to understand whether the available data from the BiB4All cohort can be used to address a local research priority. The workshop discussions also guided the topics addressed in the qualitative study in Chapter 6.

Chapter 5: Addressing local research priorities using data from the BiB4All study

5.1 Introduction

Chapter 4 identified and prioritised research topics for the BiB4All and BaBi cohort studies, as part of the first stage of the BaBi LHI Model. This chapter explores whether local research priorities can be addressed using data from the BiB4All study, which corresponds to the second stage of the BaBi LHI model (shown in Figure 15).

Figure 15 BaBi Local Health Intelligence Model*

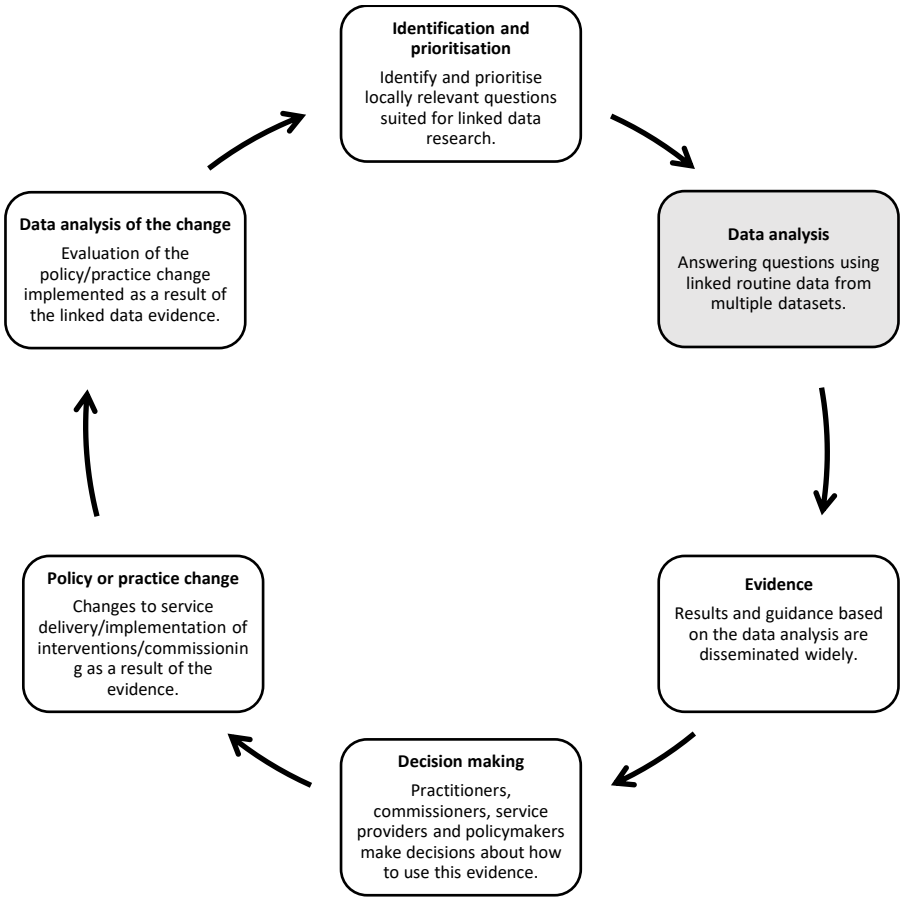


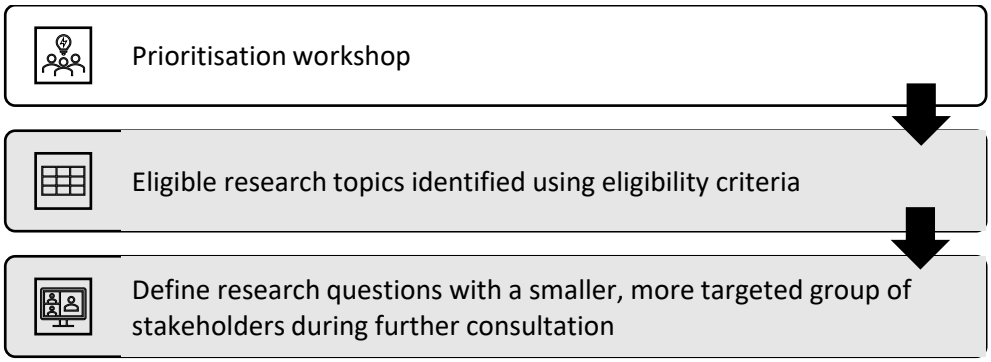
**Figure adapted from (Bryant and Bridges, 2021, Unpublished)*

Routinely collected datasets, such as those that are available as part of the BiB4All and BaBi cohorts, have been described as having many data quality issues (refer to section 1.2.2). This can present challenges for researchers utilising these data. At the time of completing this PhD, routine data from the BiB4All cohort had not been used for research. Hence, this chapter aimed to identify key challenges and opportunities

associated with using these data to address local research priorities. In addition, there was not an established process for applying for and accessing BiB4All data. Therefore, this chapter also aimed to inform how researchers apply for and use these data going forward, and to build understanding around whether these data can be used to inform decision-making. BiB4All was used as a case study as this was the originating BaBi site and was the furthest along it its development at the time this research was conducted.

In section 5.2.1 of this chapter, I utilise the research priorities identified as important and urgent in Chapter 4 and narrow these down into a list of topics that could be addressed within the scope of this thesis. Section 5.2.2 discusses how I further involved stakeholders and members of the public to decide which of these research priorities I would address using BiB4All data. This is part of the priority setting process shown in Figure 16.

Figure 16 Overview of the priority setting process for BaBi research



Section 5.3 then explores this chosen research priority using BiB4All data. It describes how these data were accessed, cleaned, and analysed, the challenges faced, and the outcomes from the data analyses. Section 5.4 discusses the findings from section 5.2 and 5.3 and provides recommendations for further research. This chapter concludes by providing four key recommendations for researchers planning to access and use BiB4All data.

5.2 Defining the research question

To decide which research question was addressed as part of the data analysis stage of the BaBi LHI model, using data from the BiB4All study, I followed these four steps:

(1) Local research priorities related to child and maternal health were identified through online prioritisation workshops (see Chapter 4).

(2) Research priorities from two of the workshops (BiB4All workshop and BaBi Doncaster workshop) were consolidated.
(3) Eligible research topics, within the context of this thesis, were identified using a criterion.
(4) The resulting research topics were defined as research questions and prioritised based on a criterion with local stakeholders.

5.2.1 Narrowing down the research priorities

As a starting point for deciding which research priority I would address with BiB4All data, I consolidated the research priorities identified in the initial BiB4All workshop (Box 11) and the BaBi Doncaster Workshop (Table 21). This allowed a broad range of topics to be included. I considered including those identified in the BaBi Wakefield workshop however, as there was a large number of attendees at that workshop, there was little time for the prioritisation task. This meant that it was difficult to determine which topics were considered most important and urgent for research. The BaBi Leeds and BaBi East London workshop outputs were also not included as they had not been hosted at the time of completing this task. Therefore, fifty-two stakeholders and members of the public from across Bradford, Leeds, Doncaster, Wakefield, and Sheffield, were involved in generating the list of research priorities that were considered for this research.

Next, I looked to narrow down this list of research priorities into those that could be addressed within the scope to this PhD, before consulting stakeholders in Bradford. The purpose of narrowing down this list before further consulting stakeholders, was to ensure I made the most out of stakeholders' time. For example, it would not be worthwhile to ask stakeholders to prioritise topics that could not be addressed with these data. Hence, I developed an eligibility criterion to identify the research priorities that met the needs to this project. This criterion is detailed in Table 24. In the BaBi toolkit, I explain how this criterion can be adapted and applied for other uses.

Table 24 Research Priority Eligibility Criterion

Inclusion	Exclusion
The question can be answered with data that are currently available, routinely collected and linked as part of the BiB4All dataset.	Requires data not routinely collected such as qualitative data.
The question has the potential to use data from more than one dataset. This means that the question could be developed to answer a broader research question.	Requires data that is not currently available such as data that will be collected in the coming years.
The question could have implications for both local policy and practice.	Research question is already being addressed by the BiB research team.
The question involves linkage to a health care dataset.	
The question has implications for child and maternal health.	

As the focus of this PhD is on the potential of using linked routine data as a local health intelligence tool, it was important that the question utilised data from multiple routinely collected data sources and had the potential to be useful to decision-makers locally. Questions that had already been explored within the BiB research team were also excluded to avoid duplication of work.

When adapting this eligibility criteria for future use, part of this adaptation could include a more formal review of the existing evidence. This is appropriate if the purpose is to inform decision-making, as researchers will want to avoid duplication of evidence. However, as the purpose of this PhD was to understand the utility of locally linked routine data to answer prioritised research questions, a more formal review of the evidence was not was not necessary. This is because the question would be used as an illustrative case study to explore the process of accessing and using locally linked routine data to address a priority area, rather than to provide an answer to this question.

Routine datasets consist of many of codes and terms that are used to record information and detailed meta-data are not always available. Hence, to understand whether the available data had the potential to address these research topics, I consulted members of the BiB research team who are familiar with linked routine data and clinicians that input these data. This allowed for a more accurate categorisation of research questions. The research priorities were also circulated to individuals' that are part of the ELP of the YHARC to determine whether any questions were already being addressed locally. Understanding existing bodies of evidence is part of the foundational phase of the FHI 360 Research Utilisation Framework (see Figure 5).

This process resulted in three eligible research priorities, which are presented in Box 12.

Box 12 Eligible research priorities

1. The effect of maternal mental health during the perinatal period on breastfeeding.
2. The effect of maternal mental health during the perinatal period on child emotional and physical development.
3. Explore the relationship between birth experience (e.g., induction of labour, assisted delivery, caesarean section) and maternal mental health.

Next, I undertook a rapid scope of the literature on the eligible research topics to understand the existing research available and whether local routine data had been used to address these topics previously. I was specifically interested in research conducted locally in Bradford or a similar community and research using routine data, as the focus of this project is around understanding the utility of locally linked routine data.

This rapid scope of the literature involved a structured search, using consistent search terms, on Google Scholar and research databases (such as Medline). I also searched the Born in Bradford website. This gave me an oversight of these topics that would allow for an informed discussion with stakeholders detailed in section 5.2.2. I established that there was limited research using locally linked routine data to explore these issues in Bradford or in a similar population. Therefore, it was appropriate for these priorities to be used to address the aims of this thesis.

I have maintained a spreadsheet detailing the prioritised research topics and the outcome of applying the eligibility criteria, which can be made available on request. It was apparent that many of the excluded research priorities would be better addressed using qualitative or bespoke research data. In agreement with the rest of the BiB4All management team, the excluded questions created a pipeline of priority projects to be dealt with when the resources became available. They also guided priority setting work as part of the YHARC.

5.2.2 Co-producing the research question

When applying the eligibility criterion, I found that the research priorities generated during the workshops were broad and needed further refinement into answerable research questions. Hence, this section describes how I further consulted local stakeholders to define the locally relevant research question that I would address using data from the BiB4All study. The term stakeholder, in this context, refers to all individuals and/or groups who have an interest in the prioritisation of child and maternal research. Example

stakeholders included parents, midwives, health visitors, early years' service commissioners, and those providing voluntary services to mothers and children.

5.2.2.1 Aim of the additional consultation

I aimed to work with local stakeholders to refine the eligible research priorities into answerable research questions and prioritise one question to be addressed as part of my PhD. This directly informed the research conducted in section 5.3.

5.2.2.2 Consultation exercise

To ensure the aim of this additional consultation was achieved, it was important that stakeholders had the opportunity to discuss the research priorities with each other. A discussion can allow stakeholders from across local early years services to share expertise on the topics, and the data collected as part of their services, to clearly define the research topics as research questions. Limited guidance is available in the literature regarding best practice for deriving research questions from broad topic areas. Hence, I developed a method of doing this, based on my experiences of conducting the workshops in Chapter 4.

Once the research questions were defined, I needed a way of prioritising one research question to be addressed as part of my PhD. Drawing on the priority setting in health care literature outlined in section 4.2.1, a range of methods were considered for this.

Criterion based approaches, and more specifically Multi-Criteria Decision Analyses (MCDA), are commonly used to prioritise health care interventions. The APEASE criteria (Affordability, Practicability, Effectiveness and Cost-effectiveness, Acceptability, Side-effects/ Safety, and Equity) is an example of a multi-criterion tool that is widely applied to design and evaluate interventions. Although this criterion is not appropriate for addressing the aims of this consultation, it offers a useful starting point for developing a criterion by which to rank the priorities for this research. As the priorities identified in Chapter 4 were considered urgent and important, incorporating a criterion for participants to consider when ranking the defined research questions could help participants reach a decision.

An adapted Delphi approach was also considered for this consultation. This would involve stakeholders attending an online meeting where they would pick their top priority based on a set of criteria. This approach was previously utilised for prioritising child

health and maternity evidence-based interventions (Forbes *et al.,* 2022), however, a reflection from this application was that decisions were weighted towards the voter's professional and personal background, and by counting votes, the outcome was biased towards how many people from a particular background attended the session. In addition, counting votes removes the discussion around why priorities were considered important. Understanding how stakeholders make their decisions can be beneficial, as it can help justify why it is important that good quality routine data on that research question is collected.

In addition, I considered an approach similar to that used as part of the JLA Priority Setting Partnership (PSP) workshops. PSP workshops apply an adapted nominal group technique where participants consider each of the topics for discussion, give their views on the topics, and vote/rank the options. This is then followed by a structured group discussion and another voting/ranking exercise. The ranking of priorities uses cards, where small groups are invited to discuss the options on the cards and organise them from most important to least important (James Lind Alliance Priority Setting Partnerships, 2022). This technique offers the opportunity to rank the priorities, with a discussion, which results in a clear top priority being identified. However, due restrictions related to the Covid-19 pandemic, this approach was not suitable. The JLA approach requires many stakeholders to be involved as they are split into small groups, it requires a long meeting, and would usually be conducted face-to-face. As health professionals had less time to engage with research due to increased clinical demands, this meant I was only able to involve a small number of stakeholders in a short meeting.

Pairwise ranking is another priority setting tool. It is becoming increasingly popular for community development projects in the field of environmental science (Russell, 1997). This approach offers the opportunity to prioritise research projects by systematically comparing each option on the list with each of the other options. In each case, one option is considered more important. This provides a more structured approach to that described as part of the JLA PSP workshops. Pairwise ranking can also allow stakeholders to apply a criterion to choose the most favourable option and drives people towards reaching a consensus, which is important for the aim of this consultation. Finally, it can allow the discussion around how people came to their decisions to be captured. Hence, the pairwise ranking approach was adopted as part of this consultation.

On the 5th April 2022, I hosted a consultation exercise on Zoom (https://zoom.us) with nine individuals from Bradford. It lasted one hour and forty-five minutes and was supported by Sally Bridges (SB). Attendees of the session represented a range of backgrounds including midwifery, health visiting, commissioner, voluntary sector, and

public contributor backgrounds. Many attendees had expertise in perinatal mental health, which was important as the eligible research priorities focused on this topic. They were able to provide valuable insights into which research priority would be most important for their services.

To invite attendees to this consultation session, I sent email invitations to those who attended the BiB4All prioritisation workshop, and who had provided their contact details to be further involvement in this research project. In addition, SB and Julie Appleyard (JA) a research midwife who is part of the BiB4All central management team, utilised their links with existing groups in Bradford and sent email invitations. I also reached out to the local early years support services using their business emails to invite a representative to attend the session. Email invitations included an information sheet and details of how to get involved. The email invitations and information sheet are provided in Appendix C. Those who responded were sent a calendar invite for the session, which included instructions on how to join the meeting.

At the start of the session, ground rules were established to ensure mutual respect and confidentiality. This was followed with a round of introductions.

Next, I delivered a short presentation about the BiB4All study and details of how the eligible research priorities were identified. I explained the purpose of the session, clarified what was meant by a research question, and described how to go about developing broad topic areas into research questions. Finally, I clarified their roles during the session and attendees had the opportunity to ask questions.

The session was split into two tasks, which were piloted with peers prior to the session. The first task asked attendees to discuss each of the research priorities in turn and clarify the exact variables of interest. The priorities were displayed on sticky notes using Google Jamboard, where I amended the sticky note based on the discussion. I then confirmed with the stakeholders that I had correctly interpreted their thoughts. I also used probes to facilitate the discussion such as:

- How would you define the perinatal period?
- What areas of child health and development are you most interested in?
- What areas of maternal mental health are most important for this research?
- Do you agree with our interpretation of birth trauma?

We also discussed the feasibility of these research questions within the limitations of routine data, drawing on their knowledge of data collected as part of their services.

Attendees were then offered a short comfort break. During this break, SB and I ensured that the discussion was accurately captured in the research questions noted on the Jamboard. After the break, I then checked attendees were happy with the questions we defined before proceeding with task two. This resulted in the following three research questions:

1. What is the effect of maternal mental health during the perinatal period on initiation and duration of breastfeeding?
2. If women experience mild to moderate mental ill health problems during the perinatal period, does this affect Ages and Stages Questionnaire outcomes at age one year and age two years?
3. Does having a different birth to your birth plan affect maternal mental health?

Task two required attendees to complete a pairwise ranking exercise. Stakeholders were asked to decide, as a group, their top priority based on a set of criteria.

As such, question one was compared to question two. If question one was considered the best option, then a '1' was placed in the corresponding cell in the pairwise ranking table: question one (row), question two (column). This was repeated until all questions had been compared to question one. This process continued for question two and three until all comparisons between questions had been made. The number of times a question was considered the top priority was measured by counting the number of times the question number had appeared in the matrix.

Box 13 sets out the criteria used to rank the priorities. This criterion was developed based on the suggestions by the CHNRI (see section 4.2.1) and through tailoring this to focus on how useful this question would be as local health intelligence. The results of task two are presented in Table 25.

Table 25 Results of pairwise ranking exercise

Question	Question Number			Score	Rank
	1	2	3		
1. The effect of mental health during the perinatal period on initiation and duration of breastfeeding.		2	1	1	2

Question	Question Number			Score	Rank
2. If women experience mild to moderate mental ill health problems during the perinatal period, does this affect Ages and Stages Questionnaire outcomes at age one and age two?			2	2	1
3.Does having a different birth to your birth plan affect maternal mental health?				0	3

Box 13 Criteria used to rank the research priorities

1. **Important for local policy and practice** – The research is important to local decision-makers and practitioners.
2. **Important for improving health locally** – is relevant to the problems faced locally in Bradford and would improve the health of the community.
3. **Feasibility of the research** – the research is realistic within the limitations of routine data, to the best of your knowledge, or would it be better addressed with another data source.
4. **Useful for decision-making** – The research produced by answering this question would be useful to policymakers. To distinguish this from criterion one, that question might be important for policy and practice but if we were to answer that question, could the evidence be used to inform decision-making. For example, the prevalence of maternal mental ill health might be important to policymakers however, what is more useful to decision-makers could be what is causing these rates of maternal mental ill health or what can prevent poor maternal mental health.
5. **Timely** – It is important that this research is addressed immediately and should not be delayed.

The pairwise ranking matrix was displayed on a Google Jamboard slide, along with a summary of the criteria, so that attendees could easily follow the process. The cells that are shaded in light grey in Table 25 are the pairwise comparisons that either compare a question with itself or are repetitions of previous comparisons noted elsewhere in the table.

The stakeholders decided the following research priority should be addressed using the BiB4All dataset as part of my PhD:

If women experience mild to moderate mental ill health problems during the perinatal period, does this affect Ages and Stages Questionnaire (ASQ) outcomes at age one and age two years?

The perinatal period was defined as the period between the point of conception to when the child is 1,001 days old. The attendees articulated that mild to moderate mental ill health was a particular concern as these women are less likely to be referred to specialist services and receive additional support, compared with women with more severe symptoms. Thus, addressing this question would support the commissioning of a service aimed at providing support to women experiencing mild to moderate mental ill health during the perinatal period. They chose the ASQ as a measure of child development, which will be described in more detail in section 5.3.

At the end of the session, attendees were asked if they could fill out a short feedback form that would be circulated after the session and thanked for their contributions. The stakeholder feedback can be found in Appendix C.

Ethical approval was not required for this stakeholder involvement; however, ethical practices were still followed. As in the prioritisation workshops, public contributors were reimbursed for their time according to the BiB payment policy for involvement in research.

5.3 Addressing the research question

This section explores whether linked routine data from the BiB4All cohort can be used to address the research priority defined by stakeholders in section 5.2. This was important as these data had not yet been accessed by researchers meaning the quality of these data were unknown. In theory, data on perinatal mental health is routinely collected by the NHS and children's ASQ scores are routinely recorded by health visitors, but are these data able to be used for research? For example, NICE guidelines recommend that the Whooley questions are completed to assess maternal mental health during routine antenatal and postnatal care, which is followed by a full mood assessment if the woman answers positively. (National Institute for Health and Care Excellence, 2014). Providers of maternity care in England are also required to submit data to NHS digital regarding whether questions for prediction and detection of mental health issues were asked during routine antenatal booking appointments. These data contribute to the national Maternity Services Data Set (NHS England, 2023). Therefore, it is expected that data regarding the Whooley questions and/or full mood assessments will be recorded routinely in the electronic health record. This section aims to explore the challenges faced in accessing

and using these data that would prevent routine data being able to inform local decision-making. Hence, this section will address two research questions:

Research Question One:
If women experience mild to moderate mental illness during the perinatal period, is this associated with the ASQ outcomes of their child at 12 months and 24 months?

As part of answering this question, this research aims to:

a) Describe the prevalence of mild to moderate mental ill health problems during the perinatal period within the cohort, and how prevalence varies according to demographic characteristics.
b) Examine whether children of women who have experienced mild to moderate mental ill health during the perinatal period have worse ASQ outcomes at 12 months and 24 months compared to children of women who have not experienced mild to moderate mental ill health during the perinatal period.
c) Examine the differences between ASQ outcomes at 12 months and 24 months for children of women who experienced mild to moderate mental ill health prenatally, compared with children of women who have experienced mild to moderate mental ill health postnatally, or throughout pregnancy.

Research Question Two:
To what extent has it been feasible to address research question one using the BiB4All linked routine data?

This section begins with an introduction to the research topic addressed as part of research question one. It outlines the relevant literature that informed the data analysis conducted in this chapter and sets out the research hypotheses. I then detail the methods and results associated with the two research questions.

5.3.1 Background on the research topic

5.3.1.1 Perinatal mental health

NHS England describe Perinatal Mental Health (PMH) problems as those which occur during pregnancy or in the first year after the child's birth, where up to 20% of new and expectant mothers are affected (NHS, 2022). Common PHM problems include perinatal depression, perinatal anxiety, perinatal Obsessive-Compulsive-Disorder (OCD), postpartum psychosis, postpartum Post-Traumatic Stress Disorder (PTSD) and eating disorders (Mind, 2022). These problems can often develop suddenly and range from mild

to severe cases (Bauer *et al.,* 2014). Research has shown that if these problems are left untreated, this can have long-lasting effects on the woman and her child (Watson *et al.,* 2019). Bauer *et al.,* (2014) reported that perinatal depression, anxiety, and psychosis carried a long-term cost of around £8.1 billion for each one-year cohort of births in the UK, and that 72% of this cost related to adverse impacts on the child rather than the mother. Risk factors for PMH problems include social, demographic, or economic disadvantage, stress, and poor social support (Prady *et al*., 2021).

Women encounter different health care services during the perinatal period including midwifery, health visiting, and primary care, which routinely collect data about them and their child. However, there are challenges associated with identifying PMH problems using routine data. In previous analysis of BiB cohort data, Prady *et al*., (2015) estimated that 50% of women experiencing Common Mental Disorders (CMD) such as anxiety and depression were potentially not diagnosed and/or treated in primary care. Willan *et al.,* (2022), also found that only 31% of women who disclosed clinically important symptoms to the BiB research team via a mental health self-reported questionnaire, had an indication of poor mental health in their primary care record. This figure increased to 42% when exploring an indication of poor mental health in both the primary care and health visitor records. A possible reason for this is that those experiencing CMD might not reveal this to health professionals. This could be due to the woman's own awareness of perinatal mental illness, the perceived stigma around disclosing a mental health problem, or the fear of being perceived as unable to parent (Khan, 2015; Insan *et al.,* 2022).

Moreover, there is consistent evidence of reduced identification and management for CMD for ethnic minority women, which is not dependant on whether they speak English (Prady *et al*., 2021; Williams *et al*., 2016; Redshaw and Henderson, 2016; Willan *et al.,* 2022). Prady *et al.* (2016) found pregnant women of Pakistani origin who were English speaking and had a lower socioeconomic status, were more likely to have an unidentified CMD, compared to those of a higher socioeconomic status. This relationship was not detected for women of other ethnic groups. Evidence suggests that health care professionals may lack the confidence to ask mothers about their mental health, especially where women come from a different culture, or when they may be time constrained during routine appointments. This creates challenges, as those who are not identified are not offered treatment or referred to specialist services, which are shown to be effective and acceptable (Evans *et al.,* 2020; Stephens *et al.,* 2016).

There are also concerns about the accuracy of PMH data. A systematic review by Davis *et al.,* (2016) explored the accuracy of administrative data from electronic health records for research into mental health diagnoses. They found that diagnoses of psychotic disorders such as schizophrenia were generally predictive of the true diagnoses but

suggested caution when interpreting the data for anxiety disorders. They suggested that routine records likely underestimate the disease burden, which is consistent with that of the other literature (Prady *et al.*, 2021 and Willan *et al.,* 2022).

Ford *et al.,* (2016) interviewed 17 UK GPs to understand how GPs document anxiety in their electronic health records. They suggested that GPs are confident in recognising anxiety symptoms but may be reluctant to code firm diagnoses in the initial stages as the condition may resolve over time. This resulted in GPs using codes that were more symptom based and details were documented in the free text fields of the medical record. This is consistent with the findings of Cresswell *et al.,* (2012) and Pybus *et al.,* (2023, Unpublished) who found that GPs frequently used the free-text field to record mental health data. Free text fields are not generally available to researchers, which creates challenges for data analysis, as routine data may provide an underestimation of the prevalence of PMH.

Prady, *et al.,* (2022, Unpublished) suggests that few clinical perinatal datasets routinely and universally capture data on social determinants other than ethnicity, age, parity, and a postcode, which allow for an area-based marker of deprivation to be derived. Issues arise when data are inputted differently according to social group. Prady, *et al.,* (2022, Unpublished) present the example, if a health care professional caring for a woman with little or no English records the mental health identification questions in the free text field but uses a clinical code when reporting this for a woman who speaks English, then it appears to researchers that the woman speaking little, or no English has not been asked those questions.

The issues highlighted here will be explored throughout this chapter and are recognised in the analyses of these data.

5.3.1.2 Perinatal mental health and child development

Maternal mental illness during the perinatal period can impact on various aspects of child development. The link between poor maternal mental health, during pregnancy and postnatally, and adverse child health outcomes is reported in a number of clinical and observational studies. Multiple studies found that poor maternal mental health during the perinatal period resulted in significant delays in offspring cognitive development (Cogill, 1986; Kiernan and Huerta, 2008; Ibanez *et al.,* 2015; Slomian *et al.,* 2019 and Kingston *et al.*, 2012). Several studies have identified a relationship between poor maternal mental health and emotional behavioural difficulties in offspring, where severity and chronicity were shown to be important factors (Giallo *et al.*, 2015; O'Connor et al., 2002; Slomian *et al.,* 2019).

There is variability in the evidence related to the relationship between maternal anxiety, depression and stress, and infant motor development. Some studies found that higher levels of prenatal anxiety, non-specific stress, and depressive symptoms are associated with more advanced motor development in children (DiPietro *et al.,*2006). However, a systematic review by Slomian *et al.,* (2019) identified studies with both significant and non-significant associations between maternal depressive symptoms and infant motor development.

A study by Mughal *et al*. (2019) examined the trajectories of maternal distress from mid-pregnancy to early childhood, on child development at three years of age, using the validated ASQ. ASQ is a screening tool used to assess child developmental progress from ages one month to five and a half years, across five domains of child development: communication, gross motor skills, fine motor skills, problem solving, personal-social development (Brookes Publishing Co., Inc, 2022). Mughal *et al*. (2019) found that children born to mothers with persistent high anxiety symptoms from pregnancy to three years postpartum had increased risk of offspring developmental delays with respect to the communication and personal-social domains. They observed a dose relationship in which the proportions of children with communication delays were highest for children of mothers reporting high anxiety symptoms, followed by subclinical symptoms and minimal symptoms. They also identified differential effects of child sex on developmental domains, where being a male child was associated with risk of delays on all domains except problem solving. They found statistically insignificant associations between maternal depression, anxiety, and stress symptoms over time and children's gross motor, fine motor, and problem-solving development at age three years.

However, some studies have found limited evidence of associations between maternal depression and adverse outcomes related to child development (Santos et al., 2010: Black et al., 2009; Brentani and Fink, 2016 and Ertel et al., 2010). Specifically, Brentani and Fink (2016) analysed the relationship between maternal depression and children's development at age one. They suggested that further research exploring whether the effect of maternal depression varies according to local context, timing, persistence, and intensity, is needed. They postulate that the lack of association could be due to the children being assessed at an average age of 12 months making it difficult to assess cognitive and socio-emotional development of the children. They also suggest that their measure of maternal mental health might not be sensitive enough to distinguish between clinically depressed and healthy mothers, leading to potential measurement error. Conflicting findings are likely the result of the diversity in the samples, methods, and measures.

Much of the research detailed in this section focuses on perinatal anxiety and depression, whereas other common mental health problems such as OCD, PTSD and eating disorders are less frequently examined. There is also less evidence related to the differences between the effects of prenatal and postnatal mental ill health and sustained mental ill health throughout the perinatal period on child development. In addition, PMH was commonly measured through questionnaires such as the Edinburgh Postnatal Depression Scale (EPDS) or through interviews with a health professional. The analysis conducted in this chapter utilises routinely collected data from multiple health datasets (including health visiting and primary care) to identify cases of mild to moderate mental ill health during the perinatal period and uses ASQ outcomes found in the health visitor record. This potentially allows a range of CMD at the mild to moderate level to be captured and the investigation into the differences in outcomes depending on whether mental ill health was experienced prenatally, postnatally, or throughout the perinatal period. This research builds on the work of Brentani and Fink (2016) by looking beyond 12 months to 24 months and explores a different local context.

5.3.2 Research Hypotheses

Based on the literature presented in section 5.3.1, I hypothesised that:

a) The prevalence of mild to moderate mental ill health during the perinatal period will differ based on ethnicity and socioeconomic status.
b) Children of women who have experienced mild to moderate mental ill health during the perinatal period have worse ASQ outcomes compared to children of women who have not experienced mild to moderate mental ill health during the perinatal period.

The evidence regarding the effects of poor maternal mental health on the five domains of child development captured by the ASQ is conflicting, therefore, I have not produced hypotheses for each domain. There are also few research studies that examine the differences in child outcomes between women who experience mental ill health prenatally, postnatally, and throughout the perinatal period, hence, I have not provided a hypothesis for this research objective.

Research question two is exploratory, therefore, I have not produced any hypotheses for this research question.

5.3.3 Methods

5.3.3.1 Data source

This study uses data from the BiB4All cohort study, which has been discussed in more detail in section 3.2.1. I utilised data from maternity, health visiting, urgent care, primary, and secondary care records, as well as pharmaceutical data.

As the recruitment for BiB4All is ongoing, this research uses data available at the date of extraction. Primary care and health visiting data were last updated on the 2nd February 2023 and the data were extracted by the BiB data team and transferred to me on the 7th February 2023. The maternity data were extracted on the 12th December 2022.

Study representativeness

To explore the representativeness of the BiB4All cohort with respect to the pregnant population living in Bradford, routine data from the Connected Bradford were used. Connected Bradford links health, education, social care, environmental, and local government data for 600,000 individuals. Connected Bradford data spans a 40-year period and is representative of the entire population of Bradford (Sohal *et al.,* 2022).

BiB4All has permission to use Connected Bradford data for the purpose of exploring how the BiB4All recruited population compares to the population of pregnant women in Bradford. This permission extends to the use of data included in this thesis.

Maternity data were extracted on the 11th June 2022 from the BiB4All and Connected Bradford cohorts. Table 26 compares the characteristics of the recruited BiB4All population of mothers to the pregnant population in Connected Bradford. The first column details the demographic characteristic of interest and the maternity data table from which these data were extracted. Table 27 compares child sex across the recruited populations of children and Table 28 compares mothers age (years) at booking between the BiB4All and Connected Bradford maternity datasets. Table 29 details how many patients were in each of the maternity data tables. The number of patients in each data table varies based on data that have been linked for consented individuals, whether a person accessed that service, and whether they had data recorded.

Due to the limitations of routine data and the scope of this project, I only compared the populations on five characteristics (maternal ethnicity, maternal age, number of pregnancies, child sex and pregnancy outcomes).

Table 26 Comparison of pregnant population in Connected Bradford to the recruited BiB4All population by variables of interest

Characteristic (dataset)	Pregnant population (Connected Bradford) N(%)	Recruited population (BiB4All) N(%)
Mother's Ethnicity (Mother Patient Table)	N=16,935	N=9,802
British (White)	5,056 (29.86%)	3,329 (33.96%)
Any other white background	1,348 (7.96%)	742 (7.57%)
White and Asian (Mixed)	82 (0.48%)	59 (0.60%)
Any other Mixed background	168 (0.99%)	85 (0.87%)
Indian (Asian or Asian British)	382 (2.26%)	163 (1.66%)
Pakistani (Asian or Asian British)	6,040 (35.67%)	3,341 (34.08%)
Bangladeshi (Asian or Asian British)	410 (2.42%)	216 (2.20%)
Any other Asian background	620 (3.66%)	424 (4.33%)
Caribbean (Black or Black British)	34 (0.20%)	18 (0.18%)
African (Black or Black British)	238 (1.41%)	120 (1.22%)
Any other Black background	76 (0.45%)	48 (0.49%)
Any other ethnic group	863 (5.09%)	433 (4.42%)
Not stated	1,618 (9.55%)	824 (8.42%)
Parity ** (Patient, Antenatal, Postnatal Table)	n=16,941	n=9652
1	4,921 (29.05%)	2,747 (28.46%)
2	4,042 (23.86%)	2,559 (26.51%)
3	2,951 (17.43%)	1,702 (17.63%)
4	2,008 (11.85%)	1,084 (11.23%)
5+	2,714 (16.00%)	1,560 (16.17%)
Unknown	305 (1.81%)	N/A
Pregnancy outcome (Patient, Antenatal, Postnatal table)	n=16,941	n=9652
Live birth	15,567 (91.89%)	8,003 (98.79%)
Miscarriage	56 (0.33%)	63 (0.78%)
Termination	23 (0.14%)	35 (0.43%)
N/A	1,295 (7.64%)	0 (0%)

**Parity includes live births, terminations and still births

Table 27 Comparison of Child Sex between pregnant populations in Connected Bradford and BiB4All

Sex (Child Patient Table)	Pregnant population (Connected Bradford) N(%) N=34,463	Recruited population (BiB4All) N(%) N=4577
Male	17,462 (50.67%)	2,340 (51.13%)
Female	16,991 (49.30%)	2,236 (48.85%)
N/A	10 (0.03%)	<5 (0.02%)

Table 28 Mother age (years) at booking compared between the BiB4All and Connected Bradford maternity datasets (Patient, Antenatal, Postnatal Table)

Population	Minimum	Maximum	Mean	Median	Standard Deviation
Eligible Population (Connected Bradford)	13	51	28.87	29	5.65
Recruited population (BiB4All)	14	48	28.8	28	5.65

Table 29 Number of patients in each data table

Table	Eligible Population (Connected Bradford)	Recruited population (BiB4All)
Mother Patient	27,135	6,381
Child Patient	34,463	4,577
Patient, Antenatal and Postnatal	16,941	9,803

Table 26, Table 27, and Table 28 demonstrate that the BiB4All population is representative of key characteristics of the pregnant population in Connected Bradford, including maternal age, ethnicity and number of pregnancies, child sex, and pregnancy outcomes.

5.3.3.2 Accessing BiB4All data: ethical considerations

Data collection

A Scientific Steering Group and Executive group review and approve studies conducted using BiB data, including BiB4All. BiB4All received ethical approval from Bradford Leeds NHS Research Ethics Committee (15/YH/0202) and the HRA. Research governance approval and sponsorship was provided by Bradford Teaching Hospitals NHS Foundation Trust. The study was also considered by the BiB Community Research Advisory Group, a Public Patient Involvement and Engagement group who discuss and contribute to the design of studies in the BiB programme.

Thus, the use of data from BiB4All is covered by its own ethics procedures. I did not seek any additional ethical approval for this study, which consists of analysis of secondary, pseudonymised data.

Data protection

All BiB data received by collaborators are pseudonymised, where identification numbers are used instead of names. BiB informs its participants that they will not be identified in any published reports (Born in Bradford 2020).

Although data are pseudonymised, Goldacre and Morely (2022) argue that participants may still be identifiable; there is an enhanced risk for pregnant women as it is possible to work out when they were in hospital based on when their child was born. Thus, data I requested are considered sensitive. To reduce the risk of identification, minimum cell counts in tables and graphs has been applied. In addition, the BiB data team generated variables that allowed me to use these data without accessing data that could potentially identify a participant.

Data Management

BiB4All data were obtained through submitting a research proposal to the BiB Executive Group. My data request was approved in May 2022 (reference number: SP612) and all data were received by February 2023. Data were provided in STATA format. A detailed list of all the codes and terms I applied for can be found in Appendix C.

Received data were stored on my personal University of York drive, accessible only by a password. I also produced a file which documented how I cleaned and analysed these data, that can be used by other researchers. This document contains decisions regarding data analysis, generation of new variables, and removing of duplicates. This file will be available to BiB when these data are returned at the end of the project, or sooner if requested.

5.3.3.3 Data choices

To identify CTV-3 codes (which are codes used to record some health data), that could be used to generate the variables needed for this analysis, I used OpenSAFELY Codelists (https://www.opencodeslists.org), an open platform for creating and sharing code lists of clinical terms and drugs. Once I had identified potentially relevant codes, I sent these to the BiB data team. They provided me with data extracts of these codes, and I determined whether they were suitable for generating the necessary variables. Often, CTV-3 codes that appeared relevant from their description, were not able to be used once I explored the data extracts. For example, there may have been a high amount of missing data, where alternative codes are used in clinical practice. Thus, the process of applying for these data was iterative, where I applied for codes, explored the data extracts, and revised the list of codes needed for the analysis.

To identify data required from the maternity datasets, the BiB data team provided files which described the Bradford Teaching Hospitals NHS Foundation Trust maternity data warehouse, including the available tables and column names. This allowed me to explore the types of data that were available, although, there were no descriptions of what these data looked like.

The process of deciding which data to apply for also involved consulting clinicians to understand how they use CTV-3 codes to record information in clinical practice. I was unable to consult any health visitors due to increased pressures on this service at the time of completing this research. Instead, I referred to a document by the Bradford District Care Trust which explained how codes are typically used by clinicians. Thus, I made assumptions about what each code or term meant, and this had implications on how I used health visiting data.

5.3.3.4 Variable choices

Variables were chosen based on previous evidence surrounding potential confounding factors in the relationship between maternal mental health and early child development outcomes. The outcome and exposure variables were chosen by local stakeholders as described in section 5.2. I sought clinical expertise from a Neonatologist and Clinical Psychologist specialising in infant and maternal health in Bradford on important variables to include in this analysis and key issues faced in Bradford. I also consulted data experts in the BiB Better Start Innovation Hub who have recently conducted research into PMH using routine data. It was important to consult clinicians throughout the project, especially

when developing variables from routine data, to ensure that I was not making incorrect assumptions about the data that are recorded.

Perinatal mental health (Exposure)

The exposure variable of interest is mild to moderate maternal mental ill health during the perinatal period.

Identifying women with mild to moderate mental ill health required a pragmatic approach to deal with the limitations of routine health data. There are many ways to record mental ill health in the routine record, where some measures are more complete than others. Recording of mental health issues can also vary between clinicians and organisations. For example, GPs may report mental health problems differently to how health visitors report these issues. In some data systems, it is only possible to record whether a questionnaire was asked but not the outcome of the questionnaire. Moreover, a woman may be diagnosed with a mental health problem by her midwife, which is recorded in the maternity dataset, but this may not be recorded in her primary care record. This creates challenges when using these data for research into PMH.

As a result, the BiB Better Start Team, led by Josie Dickerson, developed a composite indicator of mild to moderate PMH. This composite indicator is a binary variable that flags all women who have an indication of mild to moderate PMH in their routine health records, by bringing together information from maternity, health visiting and primary care datasets. To create this composite indicator, a combination of terms and CTV-3 codes relevant to CMD during the perinatal period are searched in routine health records. These terms and CTV-3 codes include signs, symptoms, diagnoses, treatments, referrals, follow up and screening (Willan *et al.,* 2022). Codes related to severe mental illness (psychoses, bipolar disorder, schizophrenia) are excluded, henceforth, I refer to this composite indicator as the PMH indicator.

The PMH indicator was developed by mapping the PMH pathway to see at what points data should be expected, this helped to ensure that the PMH indicator captured women experiencing mild forms of mental ill health. For example, the mapping process identified that health visitors offer a listening visit for women experiencing mild to moderate mental ill health, which is coded in the data and can, therefore, be included in the PMH indicator. This would have been missed if only codes related to mental health were included in the indicator. This mapping process involved a review of NICE guidance, NHS trust documentation, and consultation with local services (Willan *et al.,* 2022).

The PMH indicator was validated using the BiBBS cohort, where it was found to underestimate the prevalence of PMH conditions, although, the report does not clarify the degree of the underestimate (Willan *et al.,* 2022). This is a known limitation of using routine data to identify PMH, however, without access to other data sources, this is the best available. Willan *et al.,* (2022) suggested that the woman's symptoms of mental ill health would need to be severe in order to be recorded as clinically important. Additional research is needed to understand whether this indicator is the most useful way of identifying PMH issues using routine data. Thus, when interpreting the findings of this research, I considered these limitations and exercised caution when drawing conclusions. Despite these limitations, the focus on CMD and exclusion of codes relating to severe mental illness makes this an appropriate proxy for exploring how mild to moderate mental ill health is routinely recorded and whether this can be used for research.

Unfortunately, at the time of completing this research, the BiB data team were unable to search the terms in the maternity dataset. Hence, the PMH indicator used to identify women with mild and moderate PMH in this research was composed of codes recorded in the primary care and health visiting datasets as well as terms related to prescriptions. This means that this analysis may underestimate the true prevalence of PMH in the BiB4All sample population and some women may be categorised as experiencing no PMH if their only indication of poor mental health is recorded in their maternity record. It is also possible that this measure of PMH is too sensitive and captures women who are not experiencing PMH, as there are a number of generic mental health codes that could be used for a number of purposes. A better understanding of how these codes are used within local practice and between services is needed. However, at the time this research was completed, this indicator offered the only practical way of utilising routine data to explore PMH.

The codes used to identify women experiencing PMH in this research are available on request.

In addition, the BiB data team defined a variable which indicated whether a woman experienced mild to moderate mental ill health prenatally (from the date of conception up until the child's birthday) and postnatally (from the child's birthday up to 1,001 days after the child's birth). This was created by the data team, as this involved identifiable data. I then used this information to define the variable PMH_timing:

PMH_timing =0 if no PMH issues were experienced

PMH_timing =1 if PMH issues were experienced prenatally

PMH_timing =2 if PMH issues were experienced postnatally

PMH_timing =3 if PMH issues were experienced throughout the perinatal period

Ages and Stages Questionnaire (ASQ) (Outcome)

The primary outcome of this study is child development, measured by the Ages and Stages Questionnaire Third edition (ASQ-3). I considered exploring the Ages and Stages Questionnaire: social-emotional heath-2 (ASQ: SE-2) however, ASQ-3 was chosen due to data availability. Stakeholders were interested in ASQ outcomes at 12 and 24 months. Unfortunately, due to data availability, I was only able to look at ASQ-3 scores for 24, 27 and 30 months. I chose to utilise all three questionnaires as there was little data available at each time point.

The ASQ-3 is a parent-reported screening tool that is used in clinical and research settings to measure child development across five domains: communication, gross motor skills, fine motor skills, problem solving, personal-social development (Mughal *et al.*, 2019). It is routinely collected for children in the UK during health visiting appointments, where a score is assigned for each domain of the questionnaire (Department of Health, 2022). These ASQ assessments usually take place between nine and twelve months and 24 – 30 months and there is a questionnaire for each age range. Each questionnaire has a validation window which advises on which questionnaire should be completed depending on the child's age.

The ASQ-3 consists of 30 age-appropriate questions relating to a child's skill, ability, or behaviour. For each question, a parent can pick from one of three categories: yes (10 points), sometimes (5 points), and not yet (0 points). The scores for each item are totalled for each domain to create an overall score, where higher scores indicate more optimal development. The child's score for each domain is recorded next to a statistically derived clinical cut-off: (a) above cut-off (child is developing typically); (b) monitoring zone (a score of one standard deviation below the mean and another screening may be desirable) and (c) referral zone (child is at risk of developmental delays with a score two standard deviations below the mean and should be referred for further assessment).

For this study, I collapsed the monitoring zone and the referral zone so that the child is either categorised as typical development or at risk of developmental delays, as did Mughal *et al.*, (2019). This means risk of delay for each domain is defined as a score one standard deviation below the mean of ASQ-3 normative data. This is appropriate as this

research is concerned with understanding whether the child requires professional support, which is important for local decision-makers when allocating resources.

A dichotomous variable is more straightforward for clinicians to interpret, and they use these same cut-offs when allocating professional support. Using a clinical cut-off value also minimises the chance of identifying a false correlation, as the cut off value for *'at risk'* considers the age the questionnaire was completed. For example, the 12-month ASQ-3 can be completed for infants between 11 months and 12 months and 30 days. Infants at 11 months likely score lower on average using the ASQ-3 than infants at 12 months and 30 days, as they are two months younger and have had less time to develop. Hence, using a continuous variable to model ASQ scores can result in a false correlation as infants completing the questionnaire at 11 months are compared with those completing the questionnaire at 12 months and 30 days, where younger infants could have scored lower due to their age. Applying a cut-off value minimises the effects of age on the outcome as even if infants at 11 months score marginally lower on the ASQ due to being younger, they can still be categorised as typical development. Table 30 details the values used to categorise the ASQ-3 data.

Table 30 ASQ-3 cut-off values

ASQ Questionnaire	Domain	Cut-off value
ASQ -3 24 months	Problem Solving	39.59
	Gross motor	46.4
	Communication	38.2
	Personal-Social	41.34
	Fine Motor	43.43
ASQ-3 27 months	Problem Solving	38.79
	Gross motor	39.14
	Communication	37.22
	Personal-Social	36.11
	Fine Motor	31.08
ASQ-3 30 months	Problem Solving	38.63
	Gross motor	44.84
	Communication	43.56
	Personal-Social	41.94
	Fine Motor	33.02

* *These values are based on those provided by Squires et al., (2009)*

Hence, I used a binary variable for each ASQ domain which equals one if the child is at risk of developmental delays and equals zero for typical development. Children who completed the questionnaire outside the validation window were excluded from the

analyses. The CTV-3 codes related to ASQ-3 scores that were used in this analysis are detailed in Appendix C.

Covariates

Table 31 details the covariates included in analysis of research question one. This includes sociodemographic variables and potential confounding variables that are important for modelling the relationship between PMH and ASQ outcomes.

Table 31 List of covariates

Variable (variable name used in analysis)	Description of the variable and explanation of why included in the analyses	Dataset: codes/terms used to derive the variable	Explanation of how variable was derived using routine data
Mother's Socioeconomic Status (SES)	SES is a measure that combines an individual's economic and social status (Cockerham *et al.*, 2014). Women who are socially, demographically, or economically disadvantaged are more likely to experience poor mental health, through increased stress or discrimination. These disadvantaged women are also less likely to be diagnosed with mental ill health, offered treatment, and take up treatment (Prady *et al.*, 2021). Evidence from longitudinal studies has shown associations between low income and increased risk of delays in communication and problem solving in children as well as poor cognitive scores (Mughal *et al.*, 2019; Kiernan and Huerta, 2008). Two frameworks have been proposed to explain these findings: the family investment model and the family stress model. The family investment model theorises that income is positively associated with child development as it enables families to purchase resources, services, and experiences that are advantageous to a child's development (Kiernan and Huerta, 2008). Kiernan and Huerta (2008) highlight that this economic theory does not address how economic circumstances influences the quality of parenting. The family stress model postulates that low	N/A	Measures of SES are limited within routine data; therefore, a proxy measure was needed. Index of Multiple Deprivation (IMD) 2019 score used as a proxy for SES in this analysis. IMD is an official measure of relative deprivation for small areas of England. It combines information on the seven domains of deprivation (income deprivation; employment deprivation; education, skills, and training deprivation; health deprivation and disability; crime; barriers to housing and services and living environment deprivation) to produce a score. It also ranks every small area or neighbourhood in England from the most deprived area to the least deprived area and divides them into ten equal groups or deciles. Small areas in decile one fall within the most deprived 10% and those that are in decile ten fall within the least deprived 10%. IMD is based on data from the most recent time point available, where the most recent release was in 2019 (Ministry of Housing, Communities and Local Government, 2019). IMD for participants used in this analysis was generated by the BiB data team using postcode data available for the BiB4All participants.

Variable (variable name used in analysis)	Description of the variable and explanation of why included in the analyses	Dataset: codes/terms used to derive the variable	Explanation of how variable was derived using routine data
	income is associated with a child's development through its impact on parental mental health, which influences parenting practices and subsequently influences child outcomes. The mental health of parents and their economic circumstances are not independent, which creates challenges for research into child wellbeing (Kiernan and Huerta, 2008). Thus, I hypothesise that SES will be a confounder in the relationship between PMH and child development measured by the ASQ.		
Maternal age at child's birth (MATERNAL AGE)	Several studies have shown that maternal age is associated with child development (Falster *et al.,* 2018; Sutcliffe *et al.,* 2012; Moreno-Giménez et al., 2021). Research has also demonstrated a link between maternal age and maternal mental ill health. For example, Muraca and Joseph (2014) found that the prevalence of depression was significantly higher in women aged 40-44 years who had just delivered a baby than women aged 30–35 years who had just delivered a baby. Public Health England also suggests that young mothers up to the age of 25 are at particular risk of poor mental health (Public Health England, 2021b). Hodgkinson *et al.,* (2014) describes how adolescent and teenage mothers not only need to navigate developmental tasks but also need to adjust the responsibilities and demands of parenting, sometimes in the context of economic and social disadvantage. These additional stresses can contribute to a range of mental health problems, which can adversely affect their	Maternity data warehouse, birth table: agey_mother	Maternal age at birth is included as a categorical variable with the following categories: MATERNAL AGE= 1 if <24 years MATERNAL AGE= 0 if 25-34 years MATERNAL AGE= 2 if >35 years

Variable (variable name used in analysis)	Description of the variable and explanation of why included in the analyses	Dataset: codes/terms used to derive the variable	Explanation of how variable was derived using routine data
	parenting behaviour. There is evidence to suggest that adolescent mothers experience significantly higher rates of depression both prenatally and postnatally than adult mothers and their non-pregnant peers (Hodgkinson *et al.,* 2010). Hence, maternal age could confound the relationship between PMH and ASQ scores.		
Neonatal Intensive Care Unit Admission (NICU)	The NICU provides specialised care for sick or preterm newborn babies (Newcastle Hospitals NHS Foundation Trust. (2023). Discussions with a neonatologist revealed that removing a newborn from its mother can have profound impacts on the mother's mental health. Being admitted to the NICU can also indicate that an infant has a serious health condition or is very low birth weight, which have the potential to impact the child's development. Hence, NICU admission could be a potential confounding variable.	Maternity data warehouse, Critical Care Neonatal table: critical_care_length_of_stay	NICU admission is a binary variable which =1 if the infant was admitted to the NICU for more than one day and =0 otherwise.
Gestational age (GA)	GA is defined in weeks and is the duration of pregnancy until birth (Mizrahi, 2014). Research has shown that GA can be associated with child development (Gleason *et al.,* 2021). Babies who are born prematurely (<37 weeks' gestation) have increased risk of developmental delay (McGowan *et al.*, 2011). This risk increases for extremely premature infants (Glass *et al.*, 2015). Bowe *et al.,* (2022) also found that GA was a predictor of ASQ scores. Henderson *et al.,* (2016) found that women who have preterm births are at increased risk of ill health including stress and anxiety. This suggests that GA could be a confounder in the	Maternity data warehouse, delivery table: Gest_weeks_labour	GA at birth is measured in completed in weeks using the following categories: GA =1 if <32 weeks GA =2 if 32-33 weeks GA =3 if 34-36 weeks GA =4 if 37-38 weeks GA =0 if 39-41 weeks GA =5 if>42 weeks These categories were chosen as a result of existing literature and advice from clinical specialists and were used for descriptive statistics.

Variable (variable name used in analysis)	Description of the variable and explanation of why included in the analyses	Dataset: codes/terms used to derive the variable	Explanation of how variable was derived using routine data
	relationship between postnatal maternal mental ill health and child development. Low GA could be linked with NICU admission, however a neonatologist in Bradford advised that the majority of preterm births born at 36 weeks do not get admitted to the NICU. Based on clinical advice, GA was included in the models.		In regression analyses these categories were collapsed due to small sample sizes. The following categories were used: GA_2=1 if <37 weeks GA_2=0 if≥37 weeks
Mode of delivery (MODE OF DELIVERY)	There is evidence to suggest that the mode of delivery during childbirth is associated with maternal wellbeing following childbirth (Dekel *et al.,* 2019). In addition, multiple studies have examined the relationship between caesarean section and child development. A study by Polidano *et al.,* (2017) found a negative relationship between caesarean birth and a range of cognitive outcomes in children ages four to nine years. Takacs et al., (2021) examined the association between caesarean section and child development at age four years based on ASQ-3 and Children's Behaviour Questionnaire and Strength and Difficulties Questionnaire. They found a positive association between caesarean section and problem solving in boys and that girls were rated less optimally in the gross motor domain of the ASQ-3 when born via caesarean section. However, mode of birth was not associated with behavioural outcomes. As mode of delivery was shown to impact multiple domains of the ASQ, it is a potential confounder and is included in this analysis. Although, existing	Maternity data warehouse, birth: delivery_method.	Mode of delivery was defined using the following categories for descriptive statistics: MODE OF DELIVERY = 0 if normal or cephalic vaginal delivery MODE OF DELIVERY = 1 if forceps MODE OF DELIVERY = 2 if vacuum delivery MODE OF DELIVERY = 3 if breech MODE OF DELIVERY = 4 if elective caesarean section MODE OF DELIVERY = 5 if emergency caesarean section MODE OF DELIVERY = 6 if other MODE OF DELIVERY = 7 if not documented I used the NHS data dictionary to label the values that were provided (NHS England, 2023). Due to small sample sizes, categories were collapsed in regression analyses: MODE OF DELIVERY_2 = 0 if normal or cephalic vaginal delivery

Variable (variable name used in analysis)	Description of the variable and explanation of why included in the analyses	Dataset: codes/terms used to derive the variable	Explanation of how variable was derived using routine data
	research has focused on older children therefore, it is unclear whether there is a relationship for outcomes at ages 24, 27, and 30 months.		MODE OF DELIVERY_2 = 1 if elective caesarean section MODE OF DELIVERY_2 = 2 if emergency caesarean section MODE OF DELIVERY_2 = 3 if other
Breastfeeding for 6-8 weeks (BREASTFED)	Breastfeeding duration is considered a confounder in the relationship between postnatal maternal mental ill health and child's ASQ scores. Research evidence has shown an association between the duration of breastfeeding and improved cognitive development in children (Quigley *et al.*, 2012; Krol and Grossmann, 2018). The relationship between maternal mental ill health and breastfeeding is more complex. Maternal mental ill health early in pregnancy likely influences breastfeeding postnatally. Hamdan and Tamin (2012) revealed that higher depression scores were predictive of lower rates of breastfeeding at four months. They showed that depression scores in the third trimester of pregnancy were associated with decreased exclusive breastfeeding duration, which suggests that maternal mental health is predictive of breastfeeding behaviours in mothers. Arifunhera *et al.*, (2016) and Adedinsewo *et al.*, (2014) found that maternal anxiety reduced exclusive breastfeeding and continuation of breastfeeding. A systematic review by Slomian *et al.,* (2019) found 16 studies that showed a significant negative effect of maternal depressive symptoms on breastfeeding. Mothers with depressive	Maternity data warehouse terms, birth table: baby_brstmlk_hsp_dischrge baby_brstmlk_hsp_dischrge_desc Primary care/health visiting/healthy child programme CTV3 codes (description): Y07bf (Bottle fed 8 weeks); Y07c1 (Breast and supplement fed at 8 weeks); Y5697 (Breast fed + supp. At 6 weeks); Y07c0 (Breast fed at 8 weeks); 62P1. (Breastfed); XE1SF (infant bottle fed); 62P3 (Breast-feeding with supplement); 62P4. (Breast changed to bottle feed) 62P6 (Breast feeding stopped); 62P7 (Bottle feeding started); Y2949 (infant feeding method discussed: Bottle fed); Y2948 (infant feeding method discussed: Breast fed); Y2950 (infant feeding method discussed: Breast and bottle fed); YA737 (formula milk fed)	Breastfeeding initiation data are collected by maternity staff at birth and discharge from hospital and recorded as part of women's maternity records. A child was categorised as *"any breastfeeding"* and "no breastfeeding" at discharge from hospital. I assumed that *"Maternal breast milk"* and *"Partial breast"* meant the child was breast fed at discharge and *"Oral artificial feeds"* meant they were not breastfed at hospital discharge. The value *'%'* was recorded frequently under the baby_brstmlk_hsp_dischrge term. Using the NHS data dictionary, this was assumed to mean *"no breast milk"* or *"oral artificial feeds".* Further data around breastfeeding is collected by Health Visitors (HV) when the child is between six and eight weeks old, by the child's GP, and as part of the Healthy Child Programme, which can be accessed via primary care records. To make use of these data, I also requested the child's age at the time the CTV3 code was recorded. This was only possible for codes recorded in the child's record. This allowed me to use the generic codes for breastfeeding, such as CTV3 code 62P1, to identify children as either breastfed at six to eight weeks or not. Without this,

Variable (variable name used in analysis)	Description of the variable and explanation of why included in the analyses	Dataset: codes/terms used to derive the variable	Explanation of how variable was derived using routine data
	symptoms were significantly more likely to discontinue breastfeeding, feed their children prematurely and inappropriately, be unsatisfied with their infant feeding method, experience breastfeeding problems and bottle feed compared to mothers without depressive symptoms. Breastfeeding can also have an impact on maternal mental health postnatally. Assarian *et al.,* (2014) found that mothers who were unsuccessful at breastfeeding had a greater susceptibility to depression than those who had successful breastfeeding. A systematic review by Yuen *et al*., (2022) found thirty-six studies reporting significant relationships between breastfeeding and maternal mental health outcomes. Twenty-nine studies found that breastfeeding is associated with fewer mental health symptoms, and one found it was associated with more mental health symptoms. Borra *et al.,* (2015) concluded that the effect of breastfeeding on maternal mental health is heterogenous and is mediated by both breastfeeding intentions during pregnancy and mothers' mental health during pregnancy. Hatton *et al.,* (2005) reported a significant inverse relationship between depressive symptoms and breastfeeding at six weeks postpartum. Thus, there is potentially a reciprocal relationship between breastfeeding and maternal mental health.		there would be a significant amount of missing data, as the CTV3 code for breastfed at six to eight weeks appeared infrequently in the data. Hence, a child was categorised as "any breastfeeding" and "no breastfeeding" at six to eight weeks using primary care, health visiting and Health Child Programme data. I then generated the variable BREASTFED which =1 if the child was breastfed at discharge from hospital and at six to eight weeks. BREASTFED =0 if the child had an infant feeding code recorded in their primary care, health visiting, or Healthy Child Programme data that indicated they were not breastfed. For children without any infant feeding codes in their routine record BREASTFED = missing. The variable BREASTFED was used in descriptive statistics.

Variable (variable name used in analysis)	Description of the variable and explanation of why included in the analyses	Dataset: codes/terms used to derive the variable	Explanation of how variable was derived using routine data
	Therefore, breastfeeding duration could be considered both a mediating and confounding factor in the relationship between PMH and child development.		

Factors that are potential confounders that cannot be accounted for using available routine data included: family education; social support the mother received during the perinatal period including that from their partner; and the home environment. For example, having a supportive partner can promote a woman's wellbeing, aid the recovery from mental illness and can benefit child outcomes (Goodman, 2004; Lancaster, *et al.,* 2010). This may have implications for the analysis conducted and subsequent decision-making based on the results.

Mother's Ethnicity (Stratifying Variable)

The NHS uses 17 ethnic categories to record a patient's ethnicity, which are shown in Table 32. There is also clear guidance for the NHS that ethnicity should be self-identified by the patient (Data Set Coding Notice: 02/2001), although, there is uncertainty around whether this is adhered to in practice (Scobie *et al.*, 2021).

Table 32 Ethnicity Categories in NHS datasets*

	Code	Ethnic category description
Valid ethnic group	A	British (White)
	B	Irish (White)
	C	Any other White background
	D	White and Black Caribbean (Mixed)
	E	White and Black African (Mixed)
	F	White and Asian (Mixed)
	G	Any other Mixed Background
	H	Indian (Asian or Asian British)
	J	Pakistani (Asian or Asian British)
	K	Bangladeshi (Asian or Asian British)
	L	Any other Asian background
	M	Caribbean (Black or Black British)
	N	African (Black or Black British)
	P	Any other Black background
	R	Chinese (other ethnic group)
	S	Any other ethnic group
Not stated	Z	Not stated
Not known	X	Not known (before 2013)
	99	Not known (since 2013)
	?	Missing or values not in the NHS Data Dictionary

* *Table adapted from adapted from Scobie et al., (2021).*

It is important that minority ethnic groups are not treated as a homogenous group, but as a diverse group of individuals with varying exposure to risks that result in poor mental health. There is also a complex relationship between burden of disease and socioeconomic status as there is non-equivalence of socioeconomic status across racial groups. In planning this research, I followed the recommendations of Ross *et al.,* (2020) to avoid the pitfalls associated with using ethnicity in medical research and to ensure ethnicity is treated with respect.

Ethnic minority groups in the UK have a higher burden of CMD than the majority white population and are also less likely to be detected or treated (Watson, et al., 2019; Prady *et al.,* 2016). This may be due to increased exposure to psychosocial triggers such as deprivation, social isolation and discrimination, and inequity in access to health care. Ongoing stigma and cultural expectations could also be factors that impact on PMH for ethnic minority groups (Watson, et al., 2019). Therefore, it is important to understand the different experiences of the different ethnic groups for decision-making.

The Goldacre review, which is discussed in section 1.2.5, highlighted some of the issues with ethnicity coding in routine data. For example, ethnicity is recorded as an *"event"*, where a code is recorded during a consultation, or when first registering with a service. This means that each patient may have their ethnicity recorded more than once, and in some instances these records match and in others they conflict. Goldacre and Morley (2022) use the example of a patient being recorded as South Asian in 2004 and then Bangladeshi in 2014, to show that some patients may have fine-grained ethnicity data. It is also possible that a patient moves between groups such as being recorded as South Asian at one point in time and British Asian at another time.

A report by Scobie *et al*., (2021) on ethnicity coding in English health service datasets analysed the quality of ethnicity coding in hospital and community services datasets. They found evidence of incomplete coding, inconsistent use of codes, and an excessive number of patients having their ethnicity coded as *"not known", "not stated"* or *"other",* which impedes reliable data analysis. These data quality problems were found to disproportionately affect minority patients, where minority ethnic groups were underrepresented in health data when compared with the national population. Therefore, analysis using ethnicity data will overcount some categories for ethnicity and undercount activity for minority ethnic groups (Keith *et al.,* 2022). This must be considered when interpreting the results from this analysis.

For this analysis, I used mother's ethnicity data recorded in their maternity record. This was appropriate as these data are used for BiB4All reporting and there was a low amount of missing data. Ethnicity is also recorded in the primary care dataset and was

considered for this analysis. However, when I explored these data, there were 1,000 women with two different ethnic codes recorded.

I stratified the results of each of the analyses by ethnic group, to clarify the true significance of the findings for policy and practice. I produced descriptive statistics for a range of ethnicities to decide which groups had sufficient numbers to be carried forward into the regression analysis. Table 33 details the categories used for these descriptive statistics and the NHS codes used to create these categories.

Table 33 Ethnic groups used in this analysis

Ethnic category	NHS ethnic codes used to create this category
British (White)	A
Any other white background	B, C
White and Asian (Mixed)	F
Any other Mixed background	D, E, G
Indian (Asian or Asian British)	H
Pakistani (Asian or Asian British)	J
Bangladeshi (Asian or Asian British)	K
Any other Asian background	L
Caribbean (Black or Black British)	M
African (Black or Black British)	N
Any other Black background	P
Chinese (other ethnic group)	R
Any other ethnic group	S
Not stated	Z, 99

Child Sex (Stratifying Variable)

Mughal *et al.,* (2019) demonstrated that child sex is a predictor of child development. They found that being a male child was associated with risk of delays on all ASQ domains except problem solving. Child sex is not likely to be associated with maternal mental health and is therefore not likely to be a confounding variable (Shapiro *et al.,* 2021). However, stratifying by child sex can be useful for understanding the differences in outcomes between male and female children.

5.3.3.5 Assessment and cleaning

Simplifying the data

To simplify the data for analyses, if a mother had multiple pregnancies during the study period, her first BiB4All pregnancy with a singleton birth was included.

To clean the data, I began by inspecting each dataset individually and checking for abnormalities. I identified and removed duplicate entries by Person ID. I ensured each Child ID was associated with only one Mother ID and resolved any issues by consulting the BiB data team.

In addition to the data I requested, the data team also provided a list of mother and child Participant IDs for the cohort that were requested for matching by the data provider. This allowed me to understand how many participants should have health visiting and primary care data.

There were a few instances (n=17) where a child had multiple ASQ scores for the same domain. I excluded scores that did not form a fully completed questionnaire. Where there were two scores and two completed questionnaires, I included the questionnaire scores that were closest to the age the questionnaire should be completed. For example, if a child completed the 30-month questionnaire at 29 and 30 months, I kept the record that was completed when the child was 30 months.

PMH data that were accessed as part of this research included all PMH codes that had been recorded when the mother had been pregnant. This means the dataset included information from pregnancies pre-dating their consent to the BiB4All cohort and all pregnancies following consent. For this research, it was important to isolate the PMH codes that were recorded during their pregnancy with their first BiB4All child. This is because the research question was concerned with the impact of the mother experiencing poor PMH on her child. In order to use these data, the BiB data team provided a table which included the mother's Person ID, the date of consent, expected delivery year and expected delivery month. I used this dataset to filter on BiB4All pregnancies in the PMH datasets.

Joining the data

Each dataset had a Person ID, which related to either the mother or child participant, which was used to identify individuals in each dataset and link them up.

To join the individual datasets, I used the Birth table from the maternity database as a starting point. See Appendix C for more details of what was included in each dataset. All other datasets were added to this dataset in turn. This was appropriate as the Birth dataset had Child Person ID linked to Mother Person ID, meaning all data pertaining to each mother and child dyad could be easily added.

When I initially received the data, no dataset contained both Mother and Child Person IDs, which meant that I was unable to link the mother to the child. In collaboration with data team, we found that a delivery serial number, that was the same for both the mother and child, appeared in the Birth dataset for children and the Delivery dataset for mothers. Using this delivery serial number, the data team were able to add the mother's Person ID to the Birth dataset, linking the mother to the child. This method can be used for future research concerning mother and child dyads. In the process of doing this, there were instances where two mothers had the same delivery serial, this was logged as an issue and resolved by the BiB data team.

Once data had been cleaned and joined, there were 4,781 mother and child dyads. However, not all the children included in this dataset will be expected to have ASQ-3 scores, as they under 12 months old.

5.3.3.6 Research Question One

5.3.3.6.1 Descriptive Statistics

I summarise the BiB4All linked dataset that was accessed as part of this research. I start by describing the number of participants without a matched Systm One record (health visiting and primary care data), as data on key variables including outcome and exposure variables were missing for those individuals.

Next, I summarise the outcome and exposure variables. The sample size for children with a competed ASQ-3 is provided. Histograms for each ASQ score are presented to visualise the frequency distribution of the scores with respect to each ASQ domain. The mean and standard deviation (s.d) were also described for each domain score. Counts and percentages for the binary variables categorising children as at risk of developmental delay or typical development based on their ASQ score for each domain are presented.

The prevalence of PMH issues for the BiB4All sample, as identified by the PMH indicator, is presented. I also describe the prevalence of PMH with respect to when the women experienced this (i.e., prenatally, postnatally or throughout pregnancy) as defined by the PMH_timing variable.

Finally, I describe the covariates and stratifying variables included in the statistical analysis and present the characteristics of the BiB4All sample that have completed ASQ-3 with respect to these variables.

5.3.3.6.2 Statistical Models

This section details the statistical models used to address the research objectives for research question one.

Much of the analysis described in this section uses logistic regression analysis, which is used to examine the association between one or more independent variables and a binary (dichotomous) outcome variable (Hosmer and Lemeshow, 2013). Logistic regression is appropriate as the model parameters can provide a basis for clinically meaningful estimates of effect (Hosmer and Lemeshow, 2013).

The binary variables of interest in this research are the scores for each ASQ-3 domain, which take two values: At risk of developmental delay for that domain (=1) or typical development for that domain (=0). Formally, logistic regression can be used to estimate the probability of a particular outcome given the values of the independent variables (Schober and Vetter, 2021). It is an extension of linear regression, where instead of modelling a linear relationship between an independent variable (X) and the probability of the outcome variable (Y), which would allow for predicted probabilities outside the range of 0-1, it models a linear relationship of the independent variable with the natural logarithm (ln) of the odds of the outcome variable:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X$$

Solving this equation for the probability (p), the probability has a sigmoidal relationship with the independent variable, allowing for the estimated probabilities to be constrained between zero and one (Schober and Vetter, 2021). This approach is beneficial for research as the exponentiated logistic regression slope coefficient (e^{β}) can be interpreted as an odds ratio. An odds ratio is useful as it indicates how much the odds of the specific outcome change for a one-unit increase in the independent variable (when the independent variable is continuous) or versus a reference category (when the independent variable is categorical) (Schober and Vetter, 2021).

I acknowledge that maximum likelihood estimation is known to have small sample bias, where odds ratios tend to be too large for small samples. This may be the case in this analysis, especially where n<100. Caution is also required when interpreting standard errors and significance tests (Newsom, 2021).

Objective One: Prevalence of PMH

The variables derived in section 5.3.3.4 are used to identify the number of women who experienced PMH and the number of women who do not experience any PMH, with respect to the available demographic characteristics (socioeconomic status, ethnicity, maternal age). These descriptive statistics are presented in a summary table.

Objective Two: Maternal PMH and Child ASQ Scores

Multivariate logistic regression models are used to estimate the odds ratio for the risk of developmental delays for children aged 24, 27, and 30 months, for women identified as experiencing PMH problems by the PMH indicator compared with women who are not. The models are adjusted for the confounding variables available in routine data, that are detailed in Table 31. The basic model estimates can be described as follows:

$$Y_i = \beta_0 + \beta_1 PMH_i + \beta_2 SES_i + \beta_3 MATERNAL\ AGE_i + \beta_4 GESTATIONAL\ AGE_i + \beta_5 MODE\ OF\ DELIVERY_i + \beta_6 NICU_i$$

PMH is the perinatal mental ill health indicator and Y_i represents a binary variable that categorises the ASQ-3 score for each of the five domains at 24, 27 and 30 months, as either *"typical development"* ($Y_i = 0$) or *"at risk of developmental delay"* ($Y_i = 1$).

Stratified analysis is performed to examine these associations with respect to maternal ethnic background (where there are more than 100 participants from an ethnic group with an associated ASQ-3 score) and child sex (where there are more than 100 participants with ASQ-3 score). This provides additional information regarding the strength of association that can be used by policymakers to target interventions. I considered including an interaction term between PMH and ethnicity, however, due to small sample sizes, this was not feasible.

Objective Three: Differences in ASQ Scores in relation to timing of PMH

To understand if there are differences in ASQ scores depending on if the woman experienced mild to moderate mental ill health prenatally, postnatally, or throughout pregnancy. I planned to estimate the following models:

i. $$Y_i = \beta_0 + \beta_1 PMH\ PRENATAL_i + \beta_2 SES_i + \beta_3 MATERNAL\ AGE_i + \beta_4 GESTATIONAL\ AGE_i + \beta_5 MODE\ OF\ DELIVERY_i + \beta_6 NICU_i$$

ii. $$Y_i = \beta_0 + \beta_1 PMH\ POSTNATAL_i + \beta_2 SES_i + \beta_3 MATERNAL\ AGE_i + \beta_4 GESTATIONAL\ AGE_i + \beta_5 MODE\ OF\ DELIVERY_i + \beta_6 NICU_i + \beta_7 BREASTFED_i$$

iii. $$Y_i = \beta_0 + \beta_1 PMH\ THROUGOUT_i + \beta_2 SES_i + \beta_3 MATERNAL\ AGE_i + \beta_4 GESTATIONAL\ AGE_i + \beta_5 MODE\ OF\ DELIVERY_i + \beta_6 NICU_i$$

Y_i represents a binary variable that categorises the ASQ-3 score for each of the five domains at 24, 27 and 30 months, as either *"typical development"* ($Y_i = 0$) or *"at risk of developmental delay"* ($Y_i = 1$). Reference categories for the categorical confounding variables can be found in Table 31.

PMH PRENATAL equals one if the mother experiences prenatal mental ill health only and equals zero if no mental ill health, postnatal mental ill health only or both prenatal and postnatal mental ill health. Prenatal is defined as the period between conception and the birth of the baby. PMH POSTNATAL equals one if the mother experiences postnatal mental ill health only and equals zero if no depression, prenatal mental ill health only or both prenatal and postnatal mental ill health. Postnatal is defined as the period between after the baby is born up until the baby is 1,001 days. PHM THROUGHOUT equals one if the mother experiences both prenatal and postnatal mental ill health and equals zero if no mental ill health, prenatal mental ill health only and postnatal mental ill health only.

Inference Criteria

For each regression analysis, the exponentiated coefficients or odds ratios are reported with confidence intervals (CI) and corresponding p-values. Confidence intervals provide an expected range for the true odds ratio (Tenny and Hoffman, 2023). P-values of the Wald statistic for each variable are used to determine the significance of the variables specified in the models. A p-value is the probability under the null hypothesis of obtaining a result equal to, or more extreme than what was observed. It is the lowest significance level at which the null hypothesis can be rejected. All significant results are indicated with an asterisk.

All statistical tests are performed using STATA statistical software package and have been reviewed by my supervision team.

Model fit

The Hosmer-Lemeshow goodness-of-fit statistic was calculated for the estimated models to check how well the data fit the models. It compares the fitted and observed counts in population subgroups using a chi-squared test statistic (Hosmer and Lemeshow, 2013).

5.3.3.6.3 Sensitivity analyses

I recognise that dichotomising continuous variables can lead to misclassification bias, where values close to the cut-off point are more likely to be misclassified, as well as loss of statistical power (Suh *et al.*,1996; Selvin, 1996). It is unclear whether poor maternal mental health influences how the ASQ score is self-reported and classified. For example, it is unknown whether women who experience mild to moderate mental ill health during the perinatal period inflate or deflate the ASQ scores, making it difficult to determine whether the misclassification errors are systematic or non-systematic with respect to maternal mental health. Details of the sensitivity analyses conducted around the cut off values is detailed in Appendix C.

5.3.3.7 Research Question Two

Daas *et al.,* (2012) developed a data quality assessment framework for routine data, which focuses on three hyperdimensions: (1) the source of data; (2) meta-data and (3) data. This framework has been used by a number of national statistical institutions, including in the Netherlands, Sweden, and Australia. The source hyperdimension is concerned with the data sharing process, the meta-data hyperdimension relates to issues around quality and comprehensiveness of the associated meta-data and the data hyperdimension focuses on the quality of the actual data content (McLennan, 2018).

The data hyperdimension of the data quality assessment framework consists of five component dimensions of data quality, as described in Table 34.

Table 34 Components of the data hyperdimension of the data quality framework*

Component dimension	Description
Technical checks	Technical checks measure the useability of the data in the file, such as readability, data formatting and the degree to which the data content matches the accompanying meta-data.
Accuracy	This measures the extent to which data are correct and reliable. For example, exploring the frequency of implausible values.

Component dimension	Description
Completeness	This measures the degree to which a data source includes data describing a set of real-world objects. For example, identifying duplicate and missing values.
Time-related	This reflects the time between the data capture and supply of the data to the user.
Integrability	This measures the extent to which a data source is capable of being integrated in a statistical system. This includes for example the reliability of linking variables.

**Table adapted from Daas et al., (2012).*

This data quality assessment framework guided my assessment of the BiB4All data and its ability to address local research priorities. My assessment was also informed by the time and scope available for this research project. I focused on the technical checks, completeness, and time-related aspects of the data quality framework.

5.3.3.7.1 Assessment criteria

Timeliness and accessibility of BiB4All data

I discuss the challenges faced in accessing these data and how this potentially impacts the usefulness of these data for research. This includes a discussion of the time taken to receive the data as well as the useability of the data in the file.

Completeness of BiB4All data

The completeness of these data is explored by measuring the proportion of missing data for each of the specified variables. This is presented as part of the descriptive statistics in section 5.3.4.1. I comment on how missing data impacted the sample size and subsequent analyses.

5.3.4 Results

5.3.4.1 Research Question One

5.3.4.1.1 Descriptive Statistics

Out of 4,781 mother and child dyads, only 100 children and 24 mothers did not have a Systm one record linked. This meant they were missing key variables found in health

visiting and primary care data such as breastfeeding, ASQ and PMH data. Therefore, these participants were not included in the final analysis.

ASQ

Table 35 presents the number of BiB4All child participants that have a completed ASQ-3 at either 24, 27, or 30 months.

Table 35 Number of children with a completed ASQ-3

	Number of completed ASQ-3
ASQ-3 24 months	
Completed questionnaire	90
ASQ-3 27 months	
Completed questionnaire	413
ASQ-3 30 months	
Completed questionnaire	553

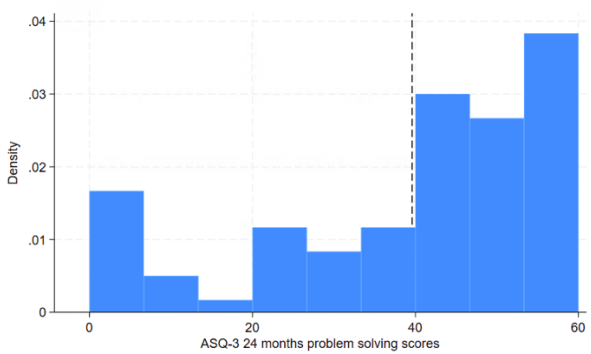
In the BiB4All dataset, there were at least 2,192 children aged 24 months or older, based on the available delivery dates, and these children were expected to have a completed ASQ-3 at 24, 27, or 30 months. Table 35 shows that ASQ-3 data were available for 1,056 children. Most of these (more than 50% of the completed ASQ-3) were recorded at 30 months. This demonstrates a high volume of missing data (>50%) related to the key outcome variable of the local research priority.

Figures 17-19 present histograms of the distribution of ASQ scores for each domain, for children with a completed ASQ-3. The vertical line represents the cut off value which categorises the child as either at risk of developmental delay or typical development. The mean and standard deviation are included below each histogram.

The histograms show that ASQ-3 scores for each domain are positively skewed, with the exception of ASQ-3 24 months communication scores, which are inversely normally distributed. Figure 17 shows that most scores were near the top score of 60 for each domain. For ASQ-3 24 months, the mean score for problem solving, gross motor skills, and fine motor scores were close to the cut off values. Hence, sensitivity analyses were conducted to see if changing the clinical cut off point by 10% affected the identified relationships. The results of this sensitivity analysis are presented in Appendix C.

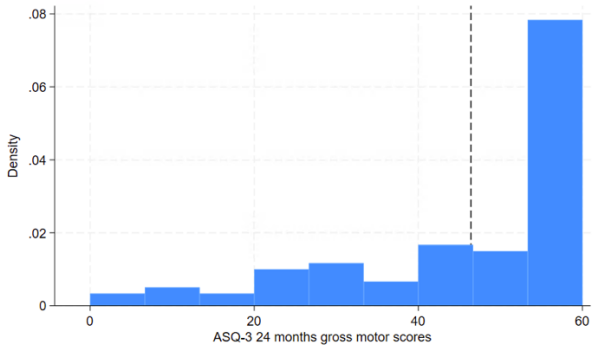
Figure 17 Histograms of ASQ-3 24-month scores for each domain

ASQ-3 24-month Problem Solving



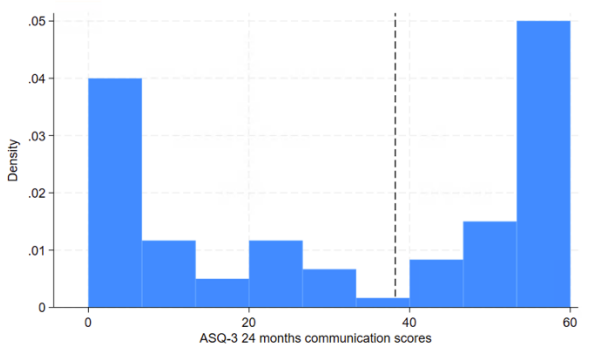
Mean score (s.d) = 39.22 (18.33)

ASQ-3 24-month Gross Motor



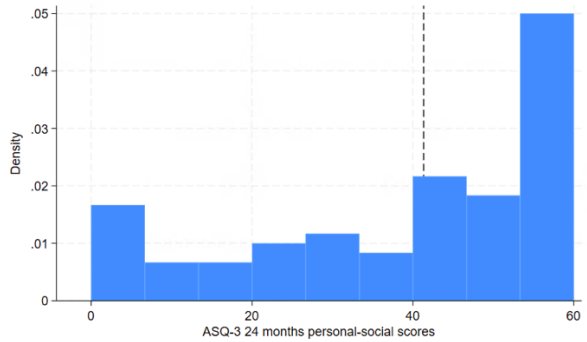
Mean score (s.d) = 46.67 (16.56)

ASQ-3 24-month Communication

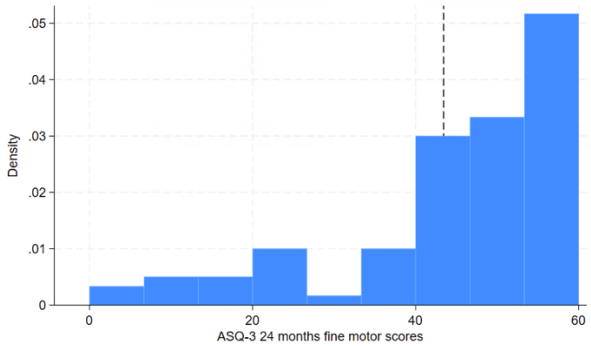


Mean score (s.d) = 32.17 (24.79)

ASQ-3 24-month Personal-Social



ASQ-3 24-month Fine Motor

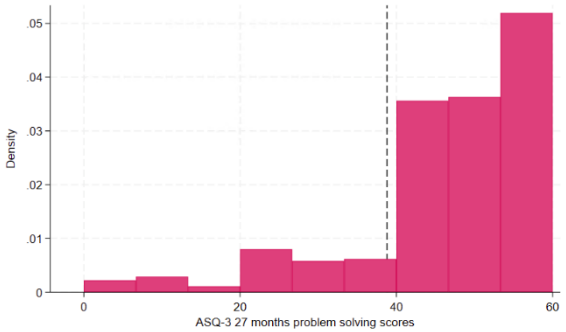


Mean score (s.d) = 38.99 (20.33)

Mean score (s.d) = 44.5 (15.26)

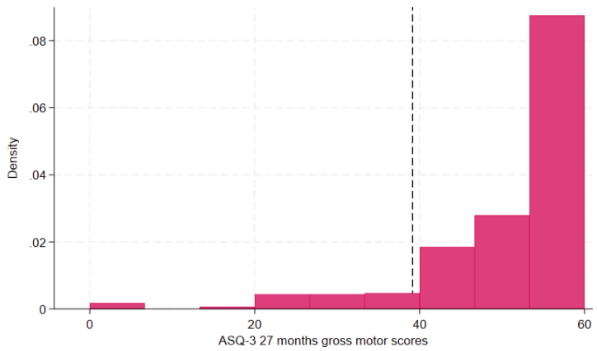
Figure 18 Histograms of ASQ-3 27-month scores for each domain

ASQ-3 27-month Problem Solving



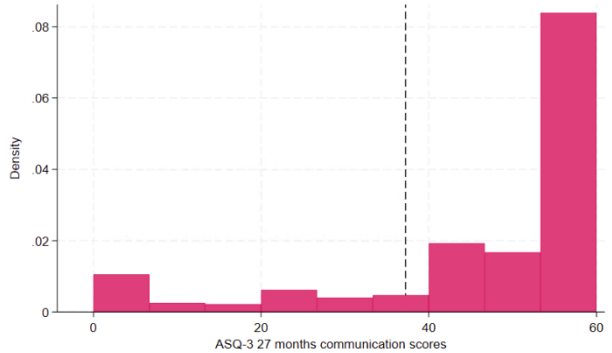
Mean score (s.d) = 46.32 (12.89)

ASQ-3 27-month Gross Motor



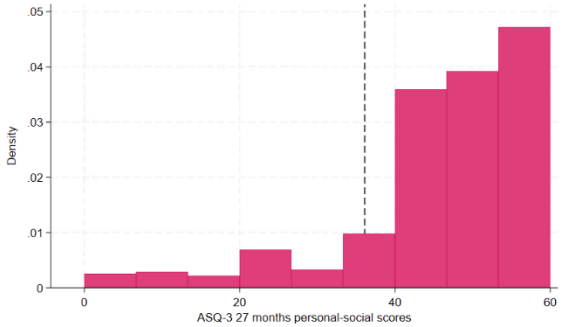
Mean score (s.d) = 51.29 (11.28)

ASQ-3 27-month Communication



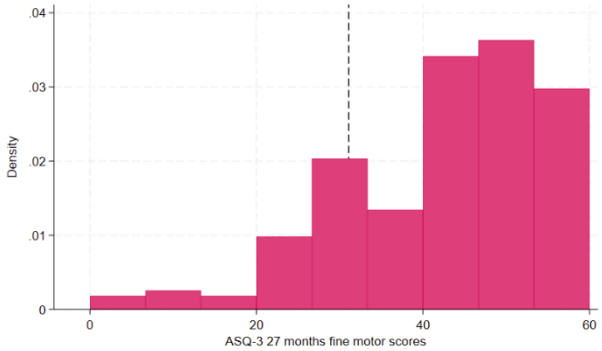
Mean score (s.d) = 47.26 (17.56)

ASQ-3 27-month Personal-Social



Mean score (s.d) = 45.74 (12.85)

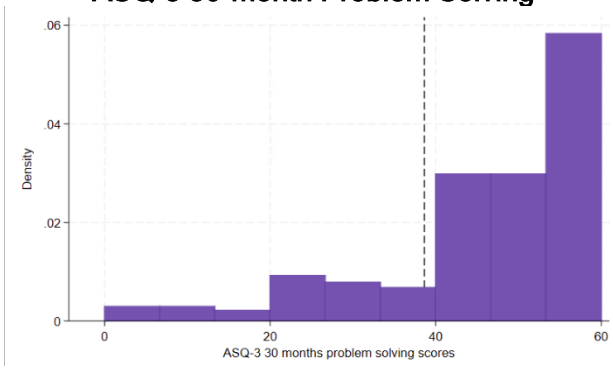
ASQ-3 27-month Fine Motor



Mean score (s.d) = 42.33 (12.82)

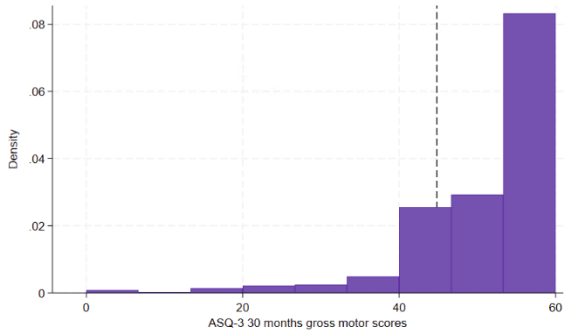
Figure 19 Histograms of ASQ-3 30-month scores for each domain

ASQ-3 30-month Problem Solving



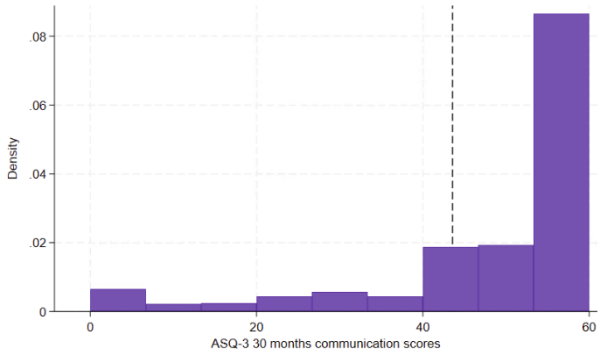
Mean score (s.d) = 45.98 (14.20)

ASQ-3 30-month Gross Motor



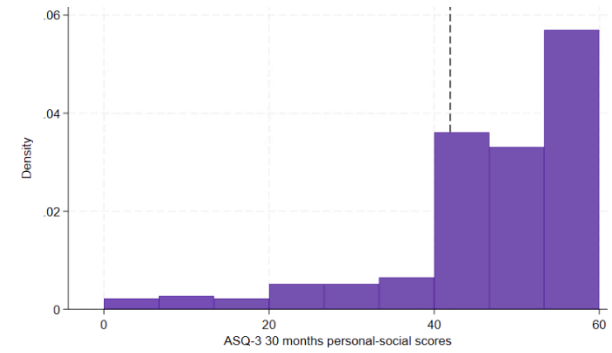
Mean score (s.d) = 51.74 (10.23)

ASQ-3 30-month Communication



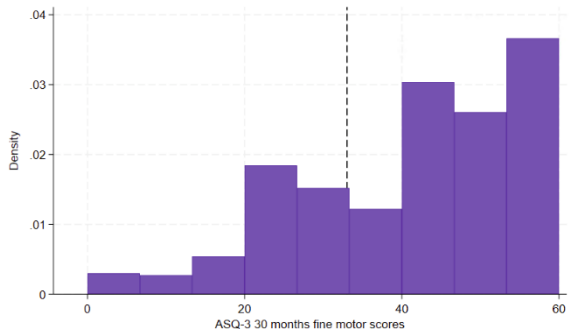
Mean score (s.d) = 49.01 (15.59)

ASQ-3 30-month Personal-Social



Mean score (s.d) = 47.01 (12.74)

ASQ-3 30-month Fine Motor



Mean score (s.d) = 40.66 (14.9)

Table 36 shows the number of children who have a completed ASQ-3 categorised as either at risk of developmental delay or typical development based on their ASQ-3 score at 24, 27 and 30 months.

Table 36 Counts and Percentages for children classified as at risk of developmental delay or typical development using ASQ-3 scores for each domain

	N (%)				
	Problem solving	**Gross motor**	**Communication**	**Personal-social**	**Fine motor**
ASQ-3 24 months (n=90)					
Typical development	57 (63.33%)	56 (62.22%)	44 (48.89%)	47 (52.22%)	58 (64.44%)
At risk of developmental delay	33 (36.67%)	34 (37.78%)	46 (51.11%)	43 (47.78)	32 (35.56%)
ASQ-3 27 months (n=413)					
Typical development	341 (82.57%)	369 (89.35%)	330 (79.90%)	337 (81.60%)	313 (75.79%)
At risk of developmental delay	72 (17.43%)	44 (10.65%)	83 (20.10%)	76 (18.40%)	100 (24.21%)
ASQ-3 30 months (n=553)					
Typical development	435 (78.66%)	468 (84.63%)	424 (76.67%)	396 (71.61%)	388 (70.16%)
At risk of developmental delay	118 (21.34%)	85 (15.37%)	129 (23.33%)	157 (28.39%)	165 (29.84%)

Table 36 shows that more than 50% of children were not at risk of developmental delay in each domain for ASQ-3 24 months, except for the communication domain where 51%

of children were at risk of developmental delay. For ASQ-3 27 and 30 months, more than 70% of children were not at risk of developmental delay for each domain.

PMH

In the BiB4All dataset, 3,359 (70.26%) women were identified as experiencing mild to moderate PMH issues using the PMH indicator. Of the women who were not identified as experiencing PMH in the routine record, 24 did not have their Systm One record linked.

Table 37 presents the number of women who experienced PMH by when they experienced the PMH (e.g., prenatally, postnatally, or throughout pregnancy).

Table 37 Counts and percentages of women experiencing PMH by timing

Timing of PMH	N (%)
No PMH	1,422 (29.74%)
Prenatally	2,396 (50.12%)
Postnatally*	12 (0.25%)
Throughout perinatal period	951 (19.89%)

**Postnatally is defined as the period between after the baby is born up until the baby is 1,001 days.*

Table 37 shows that the majority of women identified by the PMH, experienced PMH prenatally, where only a small proportion (<1%) of women experienced PMH only in the postnatal period. It also shows that 20% of women were identified as experiencing PMH both in the prenatal and postnatal period. As the number of children with a completed ASQ-3 is small, there is likely to be no variation in the outcomes of children of women identified as experiencing poor mental health only in the postnatal period.

Sample characteristics

Table 38 presents the demographic characteristics of the BiB4All sample using the available data.

Table 38 Demographic characteristics of participants included in the BiB4All linked dataset(N=4,781)

Background characteristics	N (%)	Pregnancy-specific characteristics	N (%)
Child Sex		**Neonatal admission**	
Male	2,435 (50.93%)	No	4,127 (86.32%)
Female	2,344 (49.03%)	Yes	654 (13.68%)
Missing	<10	Missing	Unknown
Mother's Ethnicity		**Mode of delivery**	
British (White)	1,773 (37.08%)	Normal or cephalic vaginal delivery	3,161 (66.12%)
Any other white background	389 (8.14%)	Forceps	336 (7.03%)
White and Asian (Mixed)	26 (0.54%)	Vacuum delivery	107 (2.24%)
Any other Mixed background	54 (1.3%)	Breech	<20
Indian (Asian or Asian British)	105 (2.20%)	Elective caesarean section	587 (12.28%)
Pakistani (Asian or Asian British)	1,469 (30.73%)	Emergency caesarean section	570 (11.92%)
Bangladeshi (Asian or Asian British)	107 (2.24%)	Other	<20
Any other Asian background	164 (3.43%)		
Caribbean (Black or Black British)	<15	**Breastfed for at least 6-8 weeks**	
African (Black or Black British)	50 (1.05%)	Yes	258 (5.4%)
Any other Black background	17 (0.36%)	No	1,998 (41.79%)
Chinese (other ethnic group)	<15	Missing	2,525 (52.81%)
Any other ethnic group	192 (4.02%)		

Background characteristics	N (%)	Pregnancy-specific characteristics	N (%)
Not stated	408 (8.53%)	**Gestational age**	
Missing	<15	<32	42 (0.88%)
		32-33	26 (0.54%)
Maternal age (years)		34-36	165 (3.45%)
<24	1,084 (22.67%)	37-38	925 (19.35%)
25-34	2,824 (59.07%)	39-41	3,015 (63.06%)
>35	873 (18.26%)	>42	68 (1.42%)
		Missing	540 (11.3%)
IMD 2019 Decile			
1	2,217 (46.37%)		
2	719 (15.04%)		
3	795 (16.63%)		
4	312 (6.53%)		
5	186 (3.89%)		
6	159 (3.32%)		
7	119 (2.49%)		
8	119 (2.49%)		
9	78 (1.63%)		
10	11 (0.23%)		
Missing	66 (1.38%)		

Data on child sex was relatively complete, with approximately 50% male children and 50% female children.

With reference to Table 38, a wide range of ethnicities are represented in the dataset, the majority of women are categorised as White British in the maternity dataset (37%, n=1773). Pakistani Asian or Asian British (30.72%, n=1469) and any other white background (8.14%, n=389) are the next largest groups. Each of the remaining groups represent <16%, with Caribbean (Black or Black British) heritage and Chinese heritage being the smallest groups. There are >408 cases (>8.53%) where the woman's ethnicity is not stated or is missing.

In the Connected Bradford dataset, which is representative of the wider population of Bradford, 35.67% of the pregnant population are of Pakistani Asian or Asian British heritage and 29.86% of the population are White British (see section 5.3.3.1). Therefore, this linked dataset is broadly representative of the wider population with regards to ethnic background.

Table 38 shows that the majority of the women included in this linked dataset (46%) live in an area that falls within the most deprived 10%. Approximately 31% of women lived in areas that fall within the second and third most deprived. This is to be expected as Bradford is one of the most deprived areas of England (City of Bradford Metropolitan District Council, 2023).

In Table 38, only a small proportion of children (14%) had a record of being admitted to the NICU. The amount of missing data for NICU admission is unknown as to categorise children as being admitted to the NICU, we rely on whether they had a record in the NICU admission dataset. An absence of a record does not mean that they were not admitted as this may have not been reported.

Most women in the sample were recorded as giving birth through normal or cephalic vaginal delivery (66%) or by caesarean section (24%). A small number of women (<10%) gave birth via a different mode of delivery.

There was a high volume (53%) of missing breastfeeding data. This means a woman's breastfeeding status at six to eight weeks was not recorded in their routine health record using the CTV-3 codes detailed in Table 31. In the absence of these data, it is unknown whether the child was breastfed or not. Where breastfeeding data were available, the majority of children (n=1,998) were not breastfed for at least six to eight weeks.

Data in Table 38 suggests that most women were recorded as giving birth between 39 and 41 weeks' gestation (63%) or between 37-38 weeks' gestation (19.35%). Less than 5% of women were recorded as giving birth prematurely (<37 weeks). Data were missing for 11% of women.

Table 38 shows that most women in the sample were aged 25-34 years (n= 2,824, 59%). The mean maternal age is 29.06 years, with the lowest age being 15 years and the maximum age being 49 years. In Table 28, which shows the age distribution of the Connected Bradford pregnant population, the mean age is 28.87 years, the lowest age is 13 years, and the maximum age is 51 years. Hence, this sample is broadly representative of the wider pregnant population with respect to maternal age, when comparing this to Table 28.

It is likely that mode of delivery and maternal age do not have any missing data as they were recorded in the *'Birth'* dataset, to which all the other datasets were linked to (see section 5.3.3.5).

Table 39 presents the characteristics of the BiB4All participants that have a completed ASQ-3 at 24, 27, or 30 months. As the number of children with ASQ scores is low, some percentages have not been presented to protect the anonymity of participants.

Table 39 Characteristics of those with completed ASQ-3, N(%)

Characteristics	Completed ASQ-3 24 months (N=90)	Completed ASQ-3 27 months (N=413)	Completed ASQ-3 30 months (N=553)
Child Sex			
Male	47 (52.22%)	207 (50.12%)	291 (52.62%)
Female	43 (47.78%)	206 (49.88%)	262 (47.38%)
Ethnicity			
British (White)	40 (44.44%)	119 (28.81%)	257 (46.47%)
Pakistani (Asian or Asian British)	26 (28.89%)	185 (44.79%)	144 (26.04%)
Any other white background	<10	19 (4.60%)	48 (8.68%)
White and Asian (Mixed	<10	<10	<10
Any other Mixed background	<10	<10	<10
Indian (Asian or Asian British)	<10	<10	<10
Bangladeshi (Asian or Asian British)	<10	13 (3.15%)	12 (2.17%)
Any other Asian background	<10	11 (2.66%)	11 (1.99%)
Caribbean (Black or Black British)	<10	<10	<10
African (Black or Black British)	<10	<10	<10
Any other Black background	<10	<10	<10
Chinese (other ethnic group)	<10	<10	<10
Any other ethnic group	<10	13 (3.15%)	13 (2.35%)
Not stated	<10	37 (8.96%)	39 (7.05%)
IMD 2019 Decile			
1	51 (56.67%)	202 (48.91%)	213 (38.52%)
2	16 (17.78%)	60 (14.53%)	78 (14.10%)
3	<10	75 (18.16%)	109 (19.71%)
4	<10	27 (6.54%)	50 (9.04%)
5	<10	11 (2.66%)	24 (4.34%)
6	<10	13 (3.15%)	18 (3.25%)
7	<10	<10	21 (3.80%)
8	<10	<10	18 (3.25%)
9	<10	<10	<15
10	<10	<10	<15

Table 39 shows that there are more than 100 participants from White British or Pakistani heritage, that have a completed ASQ-3 at 27 and 30 months. Hence, the statistical models will be stratified by these ethnicities. Table 39 shows that the majority of BiB4All participants with a completed ASQ-3 at 24, 27, or 30 months live in an area that falls within the most deprived 10% (57%, 49%, 39% respectively). The next biggest categories

are IMD 2019 deciles two and three, which suggests that most participants included in the analyses lived in the three most deprived deciles.

5.3.4.1.2 Objective One: Prevalence of PMH by demographic characteristics

Table 40 shows the number of women who were identified, by their electronic health records and using the PMH indicator, as experiencing PMH issues compared to the women who were not identified as experiencing PMH, with respect to key demographic characteristics.

Table 40 Prevalence of PMH by demographic characteristics

Variable (n=100%)	Number of women experiencing PMH issues N (%)	Number of women not identified as experiencing PMH issues N (%)
Ethnicity (n= 4,779)	**3,358 (70.27%)**	**1,421 (29.73%)**
British (White) (n=1,773)	1,185 (66.84%)	588 (33.16%)
Any other white background (n=389)	278 (71.47%)	111 (28.53%)
White and Asian (Mixed) (n=26)	16 (61.54%)	10 (38.46%)
Any other Mixed background (n=54)	35 (64.81%)	19 (35.19%)
Indian (Asian or Asian British) (n=105)	68 (64.76%)	37 (35.24%)
Pakistani (Asian or Asian British) (n=1,469)	1,085 (73.86%)	384 (26.14%)
Bangladeshi (Asian or Asian British) (n=107)	79 (73.83%)	28 (26.17%)
Any other Asian background (n=164)	115 (70.12%)	49 (29.88%)
Caribbean (Black or Black British) (n=13)	<10	<10
African (Black or Black British) (n=50)	34 (68.00%)	16 (32.00%)
Any other Black background (n=17)	<15	<15
Chinese (other ethnic group) (n=12)	<10	<10
Any other ethnic group (n=192)	148 (77.08%)	44 (22.92%)

Variable (n=100%)	Number of women experiencing PMH issues N (%)	Number of women not identified as experiencing PMH issues N (%)
Not stated (n=408)	291 (71.32%)	117 (28.68%)
Maternal age (years) (n=4,781)	**3,359 (70.26%)**	**1,422 (29.74%)**
<24 (n=1,084)	761 (70.20%)	323 (29.80%)
25-34 (n=2,824)	1,984 (70.25%)	840 (29.75%)
>35 (n=873)	614 (70.33%)	259 (29.67%)
IMD 2019 Decile (n=4,715)	**3,314 (70.29%)**	**1,401 (29.71%)**
1 (n=2,217)	1,629 (73.48%)	588 (26.52%)
2 (n=719)	507 (70.51%)	212 (29.49%)
3 (n=795)	526 (66.16%)	269 (33.84%)
4 (n=312)	218 (69.87%)	94 (30.13%)
5 (n=186)	127 (68.28%)	59 (31.72%)
6 (n=159)	109 (68.55%)	50 (31.45%)
7 (n=119)	67 (56.30%)	52 (43.70%)
8 (n=119)	78 (65.55%)	41 (34.45%)
9 (n=78)	<50	<50
10 (n=11)	<10	<10

Table 40 shows that, on average, there is a lower percentage of women identified as experiencing poor PMH for White British ethnicity than minoritised ethnic groups such as Pakistani or Bangladeshi ethnicities. The percentage of women identified as experiencing poor PMH is highest among women from any other ethnic group (77%), although this category has a small sample size.

Outlined in Table 40, a lower percentage of women are identified as experiencing PMH in less deprived deciles compared to the percentage of women identified as experiencing poor PMH in more deprived IMD deciles. For example, 73% of women living in an area in the most deprived 10% were identified as experiencing poor PMH, compared with 63% of women living in the second least deprived IMD category, although, there are lower sample sizes for higher IMD deciles.

Maternal age appeared to be unrelated to the percentage of women identified as experiencing poor PMH as, on average, 70% of women in each age category were identified as experiencing poor PMH using the PMH indicator. This was higher than expected as the NHS estimates that 20% of expectant mothers are affected by PMH

issues (NHS, 2022). This likely reflects the limitations and complexities of using routine data and the sensitivity of the PMH indicator to identify PMH issues. This is because there are a number of generic mental health codes that are being used to record PMH and it is unclear how these are being used across the system. This could suggest that the PMH indicator is too sensitive and is identifying women who have discussed their mental health, in addition to those experiencing poor PMH. Therefore, the PMH indicator, in its current form, may not be appropriate for identifying the prevalence of PMH.

5.3.4.1.3 Objective Two: Maternal PMH and Child ASQ Scores

It is important to consider the implications of using the PMH indicator to detect PMH issues, as presented under objective one, when interpreting the findings presented in this section.

Tables 41- 43 present the logistic regression analyses estimating the odds ratio for risk of developmental delay for children aged 24, 27, and 30 months respectively, adjusted for potential confounding factors. Due to small sample sizes, there are some categories, such as IMD decile eight, where there was no variation in the outcome, i.e., all children in that category were at risk of developmental delay. In these instances, the odds ratio is equal to one and there is no confidence interval.

Table 41 Logistic regression analysis estimating the odds ratio for the risk of developmental delay for children aged 24 months

ASQ domain	Problem solving (N=81)		Gross motor (N=84)		Communication (N=84)		Personal-social (N=84)		Fine motor (N=84)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH										
Indication of PMH	3.73 (0.99 to 14.12)	0.05*	4.31 (1.05 to 17.66)	0.04*	5.04 (1.45 to 17.51)	0.01*	6.84 (1.69 to 27.63)	0.01*	2.87 (0.69 to 11.96)	0.15
IMD decile 2019										
2	0.55 (1.23 to 2.37)	0.42	0.15 (0.03 to 0.87)	0.03*	0.98 (0.24 to 3.91)	0.96	0.23 (0.50 to 1.10)	0.07	0.03 (<0.01 to 0.40)	0.01*
3	1.96 (0.30 to 12.85)	0.48	0.23 (0.03 to 1.85)	0.17	0.44 (0.59 to 3.33)	0.43	0.54 (0.07 to 4.38)	0.56	0.13 (0.01 to 1.79)	0.13
4	1.31 (0.16 to 10.84)	0.80	0.20 (0.18 to 2.32)	0.20	0.61 (0.07 to 5.16)	0.65	0.13 (0.10 to 1.63)	0.11	1.37 (0.15 to 12.82)	0.78
5	4.87 (0.48 to 49.24)	0.17	0.88 (0.07 to 11.23)	0.92	3.18 (0.31 to 32.95)	0.33	3.24 (0.29 to 36.32)	0.24	1.24 (0.10 to 15.70)	0.87
6	1.00	-	0.33 (0.02 to 5.41)	0.44	2.36 (0.13 to 43.22)	0.56	3.30 (0.16 to 70.26)	0.44	0.64 (0.03 to 15.02)	0.78
7	-	-	-	-	-	-	-	-	-	-
8	1.00	-	1.00	-	1.00	-	1.00	-	1.00	-
9	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-
Maternal age										
<24 years	1.41 (0.38 to 5.34)	0.60	1.61 (0.39 to 6.62)	0.50	2.02 (0.53 to 7.66)	0.30	3.84 (0.90 to 16.47)	0.07	4.39 (0.99 to 19.51)	0.05
>35 years	0.47 (0.08 to 2.69)	0.40	0.25 (0.04 to 1.64)	0.15	0.91 (0.18 to 4.47)	0.91	0.65 (0.12 to 3.42)	0.12	0.77 (0.12 to 5.02)	0.79

ASQ domain	Problem solving (N=81)		Gross motor (N=84)		Communication (N=84)		Personal-social (N=84)		Fine motor (N=84)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
Gestational age										
<37 weeks	0.92 (0.11 to 7.81)	0.93	1.10 (0.13 to 9.67)	0.93	0.84 (0.11 to 6.68)	0.87	0.56 (0.07 to 4.70)	0.59	0.83 (0.89 to 7.62)	0.87
Mode of delivery										
Elective caesarean section	3.99 (0.92 to 17.31)	0.07	3.58 (0.46 to 16.85)	0.11	4.51 (1.01 to 20.17)	0.05	4.67 (0.98 to 22.41)	0.05	8.59 (1.58 to 46.83)	0.01*
Emergency caesarean section	3.14 (3.14 to 3.23)	0.27	1.64 (2.11 to 12.67)	0.64	1.51 (0.20 to 11.72)	0.69	1.11 (0.13 to 9.35)	0.92	2.85 (0.35 to 23.50)	0.33
Other	3.27 (0.77 to 13.86)	0.10	10.28 (1.76 to 60.17)	0.01*	4.05 (0.81 to 20.23)	0.09	2.76 (0.53 to 14.30)	0.23	1.74 (0.33 to 9.09)	0.51
NICU										
NICU admission	3.42 (0.80 to 14.70)	0.09	4.09 (0.85 to 19.67)	0.08	4.28 (0.95 to 19.38)	0.06	5.73 (1.16 to 28.36)	0.03*	2.48 (0.48 to 12.70)	0.28

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level.
OR = Odds Ratio
CI= Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

The findings in Table 41 suggest that an indication of poor PMH is not associated with risk of developmental delays in the fine motor (OR= 2.87, CI=0.69 to 11.96, p=0.15) domain of the ASQ-3 at 24 months.

Table 41 suggests that an indication of poor PMH is associated with an increased risk of developmental delays in the gross motor (OR=4.31, CI=1.05 to 17.66, p=0.04), communication (OR=5.04, CI=1.45 to 17.51, p=0.01), and personal-social (OR=6.84, CI=1.69 to 27.63, p=0.01) domains of the ASQ-3 at 24 months. As such, children of women who experience poor PMH are 6.84 times more likely to be at risk of developmental delay in the personal-social domain of the ASQ-3 24 months. It is possible that an indication of poor PMH is also associated with an increased risk of developmental delays in the problem solving (OR= 3.73, CI=0.99 to 14.12, p=0.05) domain of the ASQ-3 at 24 months as p=0.05, however the CI of the OR overlaps 1.00.

In this analysis, there is a small sample size (n<100), and odds ratios are high, which is indicative of small sample bias. The wide confidence intervals also suggest a high level of uncertainty around these estimates. The p-values associated with the Hosmer-Lemeshow goodness-of-fit statistic suggested that the models for the gross motor (p=0.04), communication (p=0.03), personal-social (p<0.01) and fine motor (p<0.01) domains of the ASQ-3 24 months may not be a good fit, which implies that these results may not be reliable.

Table 42 Logistic regression analysis estimating the odds ratio for the risk of developmental delay for children aged 27 months

ASQ domain	Problem solving (N=392)		Gross motor (N=375)		Communication (N=386)		Personal-social (N=386)		Fine motor (N=381)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH										
Indication of PMH	1.04 (0.53 to 2.08)	0.90	5.77 (1.34 to 24.87)	0.02*	0.96 (0.50 to 1.84)	0.90	1.31 (0.65 to 2.64)	0.45	0.63 (0.35 to 1.11)	0.11
IMD decile 2019										
2	0.91 (0.42 to 1.96)	0.81	1.63 (0.66 to 4.06)	0.29	0.80 (0.38 to 1.66)	0.55	1.11 (0.52 to 2.36)	0.78	1.00 (0.51 to 1.96)	0.99
3	0.56 (0.25 to 1.25)	0.16	1.07 (0.42 to 2.70)	0.89	0.45 (0.20 to 0.98)	0.04*	0.88 (0.42 to 1.84)	0.74	0.98 (0.52 to 1.83)	0.94
4	1.19 (0.44 to 3.22)	0.73	2.17 (0.71 to 6.61)	0.17	1.11 (0.44 to 2.82)	0.83	1.02 (0.36 to 2.95)	0.96	0.95 (0.37 to 2.39)	0.98
5	0.39 (0.05 to 3.24)	0.39	1.00	-	0.29 (0.04 to 2.38)	0.25	0.99 (0.20 to 4.90)	0.99	1.00	-
6	0.33 (0.04 to 2.70)	0.30	0.90 (0.11 to 7.59)	0.92	0.52 (0.11 to 2.47)	0.41	0.83 (0.17 to 4.02)	0.82	0.21 (0.03 to 1.67)	0.14
7	1.00	-	1.00	-	1.00	-	1.00	-	1.00	-
8	0.39 (0.05 to 3.38)	0.39	2.75 (0.49 to 15.40)	0.25	0.76 (0.15 to 3.85)	0.74	3.45 (0.83 to 14.28)	0.09	0.68 (0.13 to 3.45)	0.64
9	0.97 (0.11 to 8.82)	0.98	1.00	-	1.00	-	1.00	-	0.46 (0.05 to 4.16)	0.49
10	-	-	-	-	-	-	-	-	-	-
Maternal age										
<24 years	0.60 (0.28 to 1.27)	0.18	0.94 (0.37 to 2.35)	0.89	0.54 (0.26 to 1.11)	0.09	0.61 (0.29 to 1.30)	0.20	0.53 (0.28 to 1.03)	0.06
>35 years	0.81 (0.40 to 1.64)	0.56	1.17 (0.52 to 2.63)	7.11	0.93 (0.48 to 1.78)	0.82	0.78 (0.39 to 1.58)	0.50	0.97 (0.52 to 1.79)	0.91

ASQ domain	Problem solving (N=392)		Gross motor (N=375)		Communication (N=386)		Personal-social (N=386)		Fine motor (N=381)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
Gestational age										
<37 weeks	1.31 (0.47 to 3.62)	0.60	1.66 (0.49 to 5.67)	0.42	0.75 (0.25 to 2.26)	0.60	1.30 (0.48 to 3.57)	0.61	1.33 (0.52 to 3.44)	0.55
Mode of delivery										
Elective caesarean section	1.56 (0.74 to 3.28)	0.24	2.12 (0.90 to 5.01)	0.09	1.30 (0.64 to 2.65)	0.47	1.90 (0.93 to 3.91)	0.08	1.02 (0.51 to 2.06)	0.95
Emergency caesarean section	1.16 (0.49 to 2.73)	0.74	1.44 (0.49 to 4.24)	0.51	0.86 (0.36 to 2.03)	0.73	1.51 (0.68 to 3.35)	0.31	0.85 (0.39 to 1.85)	0.69
Other	1.29 (0.55 to 2.99)	0.56	1.72 (0.63 to 4.71)	0.29	1.12 (0.50 to 2.49)	0.78	0.97 (0.40 to 2.36)	0.95	0.96 (0.44 to 2.08)	0.92
NICU										
NICU admission	1.79 (0.81 to 4.00)	0.81	0.72 (0.24 to 2.19)	0.57	1.32 (0.58 to 3.00)	0.50	1.51 (0.68 to 3.36)	0.31	0.96 (0.32 to 1.13)	0.92

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI=Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 42 indicates no statistically significant associations between an indication of poor PMH and risk of developmental delays for the problem solving (OR=1.04, CI=0.53 to 2.08, p=0.90), communication (OR= 0.96, CI= 0.50 to 1.84, p=0.90), personal-social (OR= 1.31, CI= 0.65 to 2.64, p=0.45), and fine motor (OR=0.63, CI=0.35 to 1.11, p=0.11) domains of the ASQ-3 at 27 months.

Table 42 shows that an indication of poor PMH is associated with reduced odds of risk to developmental delay for communication and fine motor domains of the ASQ-3 27 months and increased odds of risk to developmental delay for the problem solving and personal-social domains, although these estimates are not statistically significant.

Table 42 suggests that children of women identified as experiencing poor PMH have increased odds of risk to developmental delays in the gross motor (OR=5.77, CI=1.34 to 24.87, p=0.03) domain of the ASQ-3 at 27 months, compared to children of women who do not experience poor PMH. However, there is a large confidence interval, which suggests a low level of precision of the odds ratio.

Table 43 Logistic regression analysis estimating the odds ratio for the risk of developmental delay for children aged 30 months

ASQ domain	Problem solving (N=527)		Gross motor (N=525)		Communication (N=527)		Personal-social (N=527)		Fine motor (N=527)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH										
Indication of PMH	2.04 (1.19 to 3.49)	0.01*	1.71 (0.95 to 3.08)	0.08	2.29 (1.34 to 3.92)	<0.01*	1.16 (0.75 to 1.79)	0.50	1.55 (1.00 to 2.40)	0.05
IMD decile 2019										
2	1.06 (0.55 to 2.02)	0.87	0.95 (0.45 to 2.03)	0.90	1.20 (0.64 to 2.24)	0.58	1.12 (0.62 to 2.05)	0.70	0.95 (0.53 to 1.71)	0.86
3	0.91 (0.51 to 1.62)	0.75	1.11 (0.59 to 2.09)	0.74	0.88 (0.49 to 1.57)	0.67	1.26 (0.75 to 2.12)	0.39	1.08 (0.65 to 1.81)	0.76
4	0.62 (0.27 to 1.44)	0.26	0.31 (0.09 to 1.08)	0.07	0.48 (0.20 to 1.17)	0.11	0.99 (0.48 to 2.03)	0.98	1.12 (0.57 to 2.23)	0.73
5	0.83 (0.29 to 2.38)	0.72	0.71 (0.19 to 2.57)	0.60	0.80 (0.28 to 2.31)	0.68	1.10 (0.43 to 2.83)	0.84	1.13 (0.46 to 2.81)	0.79
6	0.54 (0.14 to 1.98)	0.35	0.58 (0.12 to 2.75)	0.50	0.52 (0.14 to 1.93)	0.33	0.76 (0.24 to 2.46)	0.65	0.60 (0.19 to 1.95)	0.40
7	0.30 (0.07 to 1.38)	0.12	1.45 (0.48 to 4.48)	0.51	0.48 (0.13 to 1.79)	0.28	0.98 (0.35 to 2.73)	0.97	0.74 (0.25 to 2.15)	0.58
8	0.18 (0.02 to 1.42)	0.11	0.32 (0.04 to 2.49)	0.27	0.62 (0.17 to 2.27)	0.47	1.40 (0.49 to 3.95)	0.53	0.46 (0.13 to 1.66)	0.24
9	0.95 (0.24 to 3.70)	0.94	0.82 (0.17 to 4.00)	0.80	0.93 (0.24 to 3.67)	0.92	1.54 (0.47 to 5.02)	0.47	0.43 (0.09 to 2.03)	0.29
10	2.98 (0.18 to 49.44)	0.45	1.00	-	3.26 (0.20 to 54.04)	0.41	2.97 (0.18 to 49.03)	0.45	2.25 (0.14 to 37.00)	0.56

ASQ domain	Problem solving (N=527)		Gross motor (N=525)		Communication (N=527)		Personal-social (N=527)		Fine motor (N=527)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
Maternal age										
<24 years	0.86 (0.47 to 1.56)	0.62	1.02 (0.53 to 1.96)	0.95	1.30 (0.48 to 2.29)	0.35	0.90 (0.53 to 1.53)	0.71	1.37 (0.84 to 2.25)	0.20
>35 years	1.50 (0.88 to 2.56)	0.13	1.61 (0.89 to 2.92)	0.12	1.71 (1.01 to 2.90)	0.05	1.37 (0.85 to 2.22)	0.20	1.14 (0.70 to 1.87)	0.59
Gestational age										
<37 weeks	1.87 (0.47 to 1.56)	0.17	2.10 (0.82 to 5.38)	0.12	1.51 (0.63 to 3.65)	0.36	1.70 (0.74 to 3.92)	0.21	1.04 (0.45 to 2.42)	0.93
Mode of delivery										
Elective caesarean section	1.10 (0.58 to 2.10)	0.78	0.48 (0.19 to 1.20)	0.12	1.42 (0.76 to 2.62)	0.67	0.87 (0.47 to 1.61)	0.66	1.17 (0.65 to 2.09)	0.60
Emergency caesarean section	0.78 (0.34 to 1.82)	0.57	1.67 (0.76 to 3.64)	0.20	0.67 (0.29 to 1.58)	0.36	1.25 (0.63 to 2.47)	0.53	0.78 (0.38 to 1.61)	0.50
Other	0.73 (0.34 to 1.59)	0.43	0.63 (0.25 to 1.58)	0.32	0.68 (0.31 to 1.48)	0.32	1.26 (0.68 to 2.34)	0.47	0.77 (0.40 to 1.50)	0.45
NICU										
NICU admission	1.12 (0.58 to 2.19)	0.72	1.27 (0.61 to 2.64)	0.53	1.81 (0.97 to 3.39)	0.06	1.17 (0.64 to 2.14)	0.61	1.51 (0.83 to 2.72)	0.17

ASQ domain	Problem solving (N=527)		Gross motor (N=525)		Communication (N=527)		Personal-social (N=527)		Fine motor (N=527)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI= Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 43 finds no statistically significant associations between an indication of poor PMH and risk of developmental delays for the gross motor (OR=1.71, CI=0.95 to 3.08, p=0.08), personal-social (OR=1.16, CI=0.75 to 1.79, p=0.50), and fine motor (OR=1.55, CI=1.00 to 2.40, p=0.05) domains of the ASQ-3 30 months. However, the lower bound of the confidence interval associated an indication of PMH for the fine motor domain of the ASQ-3 30 months is 1.00 and p=0.05, therefore, it is possible that an indication of poor PMH in the routine record is associated with these scores.

Table 43 indicates that an indication of poor PMH is associated with an increased risk of developmental delay in the problem solving (OR=2.04, CI=1.19 to 3.49, p=0.01) and communication (OR=2.29, CI=1.34 to 3.92, p<0.01) domains of the ASQ-3 at 30 months.

Based on the p-values for the Hosmer-Lemeshow goodness-of-fit statistics for the models presented in Table 42 and Table 43, there was little evidence to suggest the models did not fit the data.

Tables 44-48 stratify the models for White British and Pakistani ethnicities, as these ethnic groups had more than 100 participants with an ASQ-3 score. ASQ-3 gross motor skills are not presented as PMH was omitted from the model due to no variation in the outcome.

Table 44 Logistic regression analysis estimating the odds ratio for the risk of developmental delay in the problem solving and communication domains of the ASQ-3 27 months stratified by ethnicity

ASQ domain	Problem solving				Communication			
	White British (N=101)		Pakistani (Asian or Asian British) (N=171)		White British (N=102)		Pakistani (Asian or Asian British) (N=171)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH								
Indication of PMH	1.58 (0.27 to 9.47)	0.61	1.54 (0.53 to 4.47)	0.43	0.98 (0.22 to 4.37)	0.98	0.88 (0.33 to 2.31)	0.79
IMD decile 2019								
2	0.85 (0.15 to 4.91)	0.86	0.92 (0.31 to 2.72)	0.88	0.46 (0.09 to 2.40)	0.36	0.57 (0.19 to 1.70)	0.31
3	0.35 (0.07 to 1.63)	0.18	0.74 (0.24 to 2.33)	0.61	0.11 (0.16 to 0.69)	0.02*	0.49 (0.15 to 1.63)	0.25
4	0.21 (0.02 to 2.58)	0.23	9.14 (1.09 to 76.69)	0.04*	0.14 (0.01 to 1.39)	0.09	4.79 (0.68 to 34.03)	0.12
5	1.00	-	1.00	-	1.00	-	1.00	-
6	0.43 (0.04 to 4.48)	0.48	1.00	-	0.58 (0.09 to 3.71)	0.57	1.00	-
7	1.00	-	1.00	-	1.00	-	1.00	-
8	1.00	-	1.00	-	0.23 (0.02 to 3.09)	0.27	1.00	-
9	1.10 (0.09 to 13.62)	0.94	-	-	1.00	-	-	-
10	-	-	-	-	-	-	-	-
Maternal age								
<24 years	0.93 (0.25 to 3.48)	0.92	0.29 (0.06 to 1.41)	0.13	0.61 (0.16 to 2.31)	0.47	0.44 (0.12 to 1.71)	0.24
>35 years	0.57 (0.05 to 5.98)	0.64	0.88 (0.34 to 2.32)	0.80	0.52 (0.08 to 3.35)	0.50	0.77 (0.29 to 2.04)	0.61

ASQ domain	Problem solving				Communication			
	White British (N=101)		Pakistani (Asian or Asian British) (N=171)		White British (N=102)		Pakistani (Asian or Asian British) (N=171)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
Gestational age								
<37 weeks	2.54 (0.29 to 22.39)	0.40	1.05 (0.17 to 6.53)	0.96	0.13 (0.01 to 3.12)	0.21	0.41 (0.04 to 3.94)	0.44
Mode of delivery								
Elective caesarean section	0.48 (0.05 to 4.65)	0.53	3.15 (1.10 to 9.08)	0.03*	1.52 (0.30 to 7.81)	0.62	2.26 (0.79 to 6.42)	0.13
Emergency caesarean section	4.89 (0.51 to 47.11)	0.17	1.73 (0.55 to 5.46)	0.55	2.30 (0.16 to 31.93)	0.54	0.80 (0.23 to 2.79)	0.73
Other	1.72 (0.27 to 11.14)	0.57	2.10 (0.67 to 6.53)	0.20	4.26 (0.79 to 22.92)	0.09	1.39 (0.47 to 4.15)	0.55
NICU								
NICU admission	0.72 (0.08 to 6.81)	0.77	1.34 (0.44 to 4.11)	0.61	3.52 (0.39 to 32.13)	0.26	1.01 (0.31 to 3.37)	0.98

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI = Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 45 Logistic regression analysis estimating the odds ratio for the risk of developmental delay in the personal-social and fine motor domains of the ASQ-3 27 months stratified by ethnicity

ASQ domain	Personal-social				Fine motor			
	White British (N=85)		Pakistani (Asian or Asian British) (N=171)		White British (N=107)		Pakistani (Asian or Asian British) (N=171)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH								
Indication of PMH	3.58 (0.40 to 32.43)	0.26	0.91 (0.32 to 2.60)	0.87	1.28 (0.29 to 5.59)	0.74	0.43 (0.18 to 1.00)	0.05
IMD decile 2019								
2	1.77 (0.31 to 10.26)	0.52	0.54 (0.17 to 1.74)	0.30	1.13 (0.24 to 5.32)	0.87	1.07 (0.42 to 2.73)	0.89
3	0.84 (0.19 to 3.73)	0.82	0.39 (0.11 to 1.38)	0.14	0.40 (0.10 to 1.56)	0.19	1.16 (0.44 to 3.10)	0.76
4	1.00	-	3.21 (0.37 to 27.96)	0.29	0.46 (0.07 to 2.81)	0.40	0.51 (0.05 to 5.35)	0.57
5	1.00	-	1.00	-	1.00	-	1.00	-
6	1.10 (0.16 to 7.63)	0.93	1.00	-	0.21 (0.02 to 2.06)	0.18	1.00	-
7	1.00	-	1.00	-	1.00	-	1.00	-
8	2.06 (0.21 to 20.38)	0.54	1.00	-	0.51 (0.04 to 6.00)	0.59	1.00	-
9	1.00	-	-	-	0.49 (0.04 to 5.73)	0.57	-	-
10	-	-	-	-	-	-	-	-
Maternal age								
<24 years	0.30 (0.06 to 1.64)	0.17	0.32 (0.07 to 1.56)	0.16	0.73 (0.22 to 2.41)	0.61	0.22 (0.05 to 1.03)	0.06
>35 years	0.37 (0.04 to 3.68)	0.40	0.68 (0.24 to 1.91)	0.46	0.67 (0.11 to 3.95)	0.66	1.30 (0.56 to 3.02)	0.54
Gestational age								

ASQ domain	Personal-social				Fine motor			
	White British (N=85)		Pakistani (Asian or Asian British) (N=171)		White British (N=107)		Pakistani (Asian or Asian British) (N=171)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
<37 weeks	1.57 (0.15 to 16.72)	0.71	0.75 (0.11 to 5.18)	0.78	0.80 (0.07 to 9.40)	0.86	0.85 (0.16 to 4.56)	0.85
Mode of delivery								
Elective caesarean section	1.09 (0.16 to 7.65)	0.93	4.44 (1.48 to 13.34)	0.01*	0.12 (0.02 to 1.81)	0.15	2.04 (0.75 to 5.52)	0.16
Emergency caesarean section	1.00	-	3.19 (1.02 to 9.92)	0.05	0.92 (0.07 to 11.59)	0.95	0.74 (0.26 to 2.14)	0.58
Other	1.76 (0.29 to 10.61)	0.54	1.81 (0.54 to 6.09)	0.34	0.75 (0.13 to 4.23)	0.74	1.40 (0.49 to 4.04)	0.53
NICU								
NICU admission	0.90 (0.08 to 9.90)	0.93	2.01 (0.67 to 6.07)	0.22	0.38 (0.32 to 4.56)	0.45	1.61 (0.56 to 4.60)	0.37

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI = Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 44 and Table 45 show that an indication of PMH is not associated with risk to developmental delays in the problem solving (OR=1.58, CI=0.27 to 9.47, p=0.61; OR=1.54, CI=0.53 to 4.47, p=0.43), communication (OR=0.98, CI=0.22 to 4.37, p=0.98; OR=0.88 CI=0.33 to 2.31, p=0.79), personal-social (OR=3.58, CI=0.40 to 32.43, p=0.26; OR=0.91, CI=0.32 to 2.60, p=0.87), and fine motor (OR=1.28, CI=0.29 to 5.59, p=0.74; OR=0.43, CI=0.18 to 1.00, p=0.05) domains of the ASQ-3 27 months for White British and Pakistani ethnicities respectively.

The estimates are associated with wide confidence intervals, suggesting high levels of uncertainty around these values. In addition, the upper bound of the confidence interval associated with the fine motor domain of the ASQ 27 months for women of Pakistani heritage is 1.00, which could suggest that this PMH may be associated with risk to developmental delay in this domain for larger sample sizes. Furthermore, the Hosmer-Lemeshow goodness-of-fit test suggested that the model for the fine motor domain of the ASQ-3 27 months for Pakistani women may not be a good fit, suggesting the results may not be reliable.

Based on the p-values of the Hosmer-Lemeshow goodness-of-fit statistics for the other models presented in Table 44 and Table 45, there is limited evidence to suggest that the models did not fit the data.

Table 46 Logistic regression analysis estimating the odds ratio for the risk of developmental delay in the problem solving and gross motor domains of the ASQ-3 30 months stratified by ethnicity

ASQ domain	Problem solving				Gross motor			
	White British (N=235)		Pakistani (Asian or Asian British) (N=132)		White British (N=226)		Pakistani (Asian or Asian British) (N=126)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH								
Indication of PMH	2.60 (1.06 to 6.34)	0.04*	2.32 (0.90 to 6.01)	0.08	1.18 (0.53 to 2.63)	0.69	5.14 (1.30 to 20.24)	0.02*
IMD decile 2019								
2	0.61 (0.18 to 2.08)	0.43	1.63 (0.55 to 4.85)	0.38	1.95 (0.64 to 5.98)	0.24	0.51 (0.12 to 2.20)	0.37
3	0.59 (0.19 to 1.81)	0.36	1.01 (0.38 to 2.63)	0.99	1.62 (0.57 to 4.56)	0.36	0.99 (0.34 to 2.85)	0.98
4	0.49 (0.15 to 1.64)	0.25	1.30 (0.20 to 8.62)	0.789	0.61 (0.16 to 2.36)	0.48	1.00	-
5	0.68 (0.12 to 3.74)	0.66	2.45 (0.45 to 13.25)	0.30	1.00	-	2.95 (0.50 to 17.61)	0.23
6	0.38 (0.04 to 3.38)	0.39	1.91 (0.11 to 33.64)	0.66	1.00	-	2.40 (0.13 to 43.32)	0.55
7	0.23 (0.03 to 2.07)	0.19	1.00	-	2.75 (0.66 to 11.51)	0.17	1.00	-
8	1.00	-	1.00	-	0.62 (0.07 to 5.44)	0.66	1.00	-
9	2.11 (0.44 to 10.07)	0.35	-	-	1.62 (0.28 to 9.34)	0.59	-	-
10	2.92 (0.17 to 51.28)	0.46	-	-	1.00	-	-	-
Maternal age								
<24 years	1.04 (0.44 to 2.46)	0.92	0.69 (0.22 to 2.17)	0.52	1.64 (0.68 to 3.91)	0.27	0.24 (0.05 to 1.26)	0.09

ASQ domain	Problem solving				Gross motor			
	White British (N=235)		Pakistani (Asian or Asian British) (N=132)		White British (N=226)		Pakistani (Asian or Asian British) (N=126)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
>35 years	0.74 (0.25 to 2.17)	0.58	1.13 (0.45 to 2.83)	0.80	0.86 (0.29 to 2.53)	0.78	1.27 (0.44 to 3.66)	0.66
Gestational age								
<37 weeks	2.75 (0.68 to 11.18)	0.16	0.47 (0.03 to 7.38)	0.59	3.15 (0.71 to 13.88)	0.13	3.30 (0.15 to 72.79)	0.45
Mode of delivery								
Elective caesarean section	1.42 (0.48 to 4.19)	0.52	1.66 (0.47 to 5.86)	0.43	0.18 (0.23 to 1.44)	0.11	0.83 (0.17 to 4.07)	0.81
Emergency caesarean section	0.43 (0.08 to 2.23)	0.31	0.65 (0.17 to 2.48)	0.53	1.68 (0.51 to 5.52)	0.39	2.30 (0.54 to 9.82)	0.26
Other	0.53 (0.11 to 2.58)	0.43	0.74 (0.21 to 2.69)	0.65	0.28 (0.03 to 2.22)	0.23	2.05 (0.49 to 8.49)	0.32
NICU								
NICU admission	1.35 (0.43 to 4.24)	0.61	0.98 (0.28 to 3.45)	0.98	0.90 (0.26 to 3.16)	0.87	0.86 (0.21 to 3.56)	0.84

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI = Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 46 shows that having indication of poor PMH is associated with an increased odds (OR=2.60, CI=1.06 to 6.34, p=0.04) of risk to developmental delays for the problem-solving domain of the ASQ-3 30 months for children of White British mothers. There is also an increased odds (OR=2.32, CI=0.90 to 6.01, p=0.08) for risk of developmental delay in the problem-solving domain of ASQ-3 30 months for children of Pakistani mothers who have an indication of PMH, although this was not statistically significant.

Table 46 also shows that an indication of poor PMH is associated with an increased odds (OR=5.14, CI=1.30 to 20.24, p=0.02) of risk to developmental delay for the gross motor domain of the ASQ-3 30 months for children of Pakistani mothers. There are increased odds (OR=1.18, CI=0.53 to 2.63, p=0.69) for children of White British mothers who experience poor PMH in the gross motor domain of the ASQ-3 30 months, although this was not statistically significant.

The estimated odds ratios in Table 46 are associated with large confidence intervals, likely due to small sample sizes, meaning there is a high level of uncertainty around these estimates. The p-values associated with the Hosmer-Lemeshow goodness-of-fit statistic suggested there was limited evidence that the models presented Table 46 were not a good fit for the data.

Table 47 Logistic regression analysis estimating the odds ratio for the risk of developmental delay in the communication and personal-social domains of the ASQ-3 30 months stratified by ethnicity

ASQ domain	Communication				Personal Social			
	White British (N=248)		Pakistani (Asian or Asian British) (N=132)		White British (N=248)		Pakistani (Asian or Asian British) (N=134)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH								
Indication of PMH	2.41 (1.03 to 5.63)	0.04*	3.23 (1.18 to 8.90)	0.02*	2.40 (1.14 to 5.05)	0.02*	0.87 (0.38 to 2.00)	0.75
IMD decile 2019								
2	0.87 (0.28 to 2.74)	0.82	1.63 (0.53 to 5.07)	0.39	0.89 (0.29 to 2.74)	0.84	1.26 (0.43 to 3.66)	0.67
3	0.44 (0.13 to 1.47)	0.18	1.37 (0.53 to 3.59)	0.52	0.80 (0.29 to 2.21)	0.67	0.99 (0.40 to 2.46)	0.98
4	0.67 (0.22 to 2.04)	0.48	0.45 (0.04 to 4.78)	0.51	0.92 (0.33 to 2.53)	0.87	1.03 (0.16 to 6.64)	0.97
5	0.83 (0.15 to 4.53)	0.83	2.33 (0.39 to 13.97)	0.35	0.93 (0.18 to 4.86)	0.93	6.30 (0.10 to 31.58)	0.05
6	0.42 (0.05 to 3.73)	0.44	2.08 (0.12 to 36.85)	0.62	1.94 (0.42 to 8.95)	0.39	1.80 (0.10 to 31.57)	0.69
7	0.53 (0.09 to 2.94)	0.47	1.00	-	1.06 (0.25 to 4.56)	0.94	2.13 (0.10 to 45.55)	0.10
8	0.35 (0.04 to 3.01)	0.34	1.00	-	2.63 (0.73 to 9.50)	0.14	1.00	-
9	2.02 (0.43 to 9.56)	0.38	-	-	2.31 (0.53 to 10.07)	0.27	-	-
10	2.93 (0.17 to 51.21)	0.46	-	-	3.49 (0.20 to 60.84)	0.39	-	-
Maternal age								
<24 years	1.07 (0.46 to 2.49)	0.87	0.68 (0.21 to 2.20)	0.52	1.59 (0.73 to 3.46)	0.25	0.43 (0.13 to 1.39)	0.16

ASQ domain	Communication				Personal Social			
	White British (N=248)		Pakistani (Asian or Asian British) (N=132)		White British (N=248)		Pakistani (Asian or Asian British) (N=134)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
>35 years	0.78 (0.28 to 2.17)	0.64	1.12 (0.44 to 2.87)	0.81	1.41 (0.59 to 3.37)	0.44	1.17 (0.48 to 2.87)	0.73
Gestational age								
<37 weeks	3.19 (0.85 to 12.04)	0.09	0.66 (0.05 to 9.23)	0.76	1.89 (0.49 to 7.21)	0.35	0.19 (0.01 to 2.69)	0.22
Mode of delivery								
Elective caesarean section	0.89 (0.29 to 2.71)	0.83	2.40 (0.67 to 8.56)	0.18	0.64 (0.22 to 1.91)	0.43	1.17 (0.34 to 4.09)	0.80
Emergency caesarean section	0.68 (0.16 to 2.84)	0.59	0.42 (0.10 to 1.82)	0.25	2.25 (0.75 to 6.77)	0.15	0.46 (0.13 to 1.63)	0.23
Other	0.21 (0.03 to 1.74)	0.15	0.71 (0.19 to 2.65)	0.62	0.92 (0.28 to 3.06)	0.90	0.77 (0.23 to 2.59)	0.67
NICU								
NICU admission	1.39 (0.47 to 4.16)	0.55	2.81 (0.82 to 9.66)	0.10	0.80 (0.27 to 2.41)	0.70	2.68 (0.82 to 8.79)	0.10

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI = Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 48 Logistic regression analysis estimating the odds ratio for the risk of developmental delay in the fine motor domain of the ASQ-3 30 months stratified by ethnicity

ASQ domain	Fine Motor			
	White British (N=248)		Pakistani (Asian or Asian British) (N=134)	
	OR (CI)	P value	OR (CI)	P value
PMH				
Indication of PMH	2.51 (1.27 to 4.97)	0.01*	1.20 (0.50 to 2.90)	0.68
IMD decile 2019				
2	0.45 (0.16 to 1.59)	0.14	1.48 (0.48 to 4.57)	0.50
3	0.74 (0.30 to 1.79)	0.50	1.33 (0.51 to 3.45)	0.56
4	0.80 (0.32 to 1.96)	0.62	2.28 (0.36 to 14.41)	0.38
5	0.44 (0.08 to 2.27)	0.32	5.25 (0.95 to 29.13)	0.06
6	0.75 (0.17 to 3.38)	0.71	2.22 (0.12 to 39.82)	0.59
7	0.58 (0.14 to 2.40)	0.45	3.37 (0.15 to 75.40)	0.44
8	0.62 (0.15 to 2.55)	0.50	1.00	-
9	0.47 (0.09 to 2.55)	0.38	-	-
10	1.44 (0.08 to 24.71)	0.80	-	-
Maternal age				

ASQ domain	Fine Motor			
	White British (N=248)		Pakistani (Asian or Asian British) (N=134)	
	OR (CI)	P value	OR (CI)	P value
<24 years	1.31 (0.65 to 2.66)	0.46	0.74 (0.23 to 2.38)	0.61
>35 years	0.75 (0.32 to 1.79)	0.52	1.48 (0.59 to 3.71)	0.40
Gestational age				
<37 weeks	0.78 (0.21 to 2.94)	0.71	2.39 (0.16 to 36.36)	0.53
Mode of delivery				
Elective caesarean section	0.91 (0.36 to 2.33)	0.85	3.09 (0.87 to 10.98)	0.08
Emergency caesarean section	0.78 (0.25 to 2.45)	0.68	0.42 (0.11 to 1.66)	0.22
Other	0.56 (0.18 to 1.90)	0.37	0.47 (0.12 to 1.91)	0.29
NICU				
NICU admission	1.97 (0.75 to 5.16)	0.17	1.37 (0.39 to 4.78)	0.63

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI= Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 47 and Table 48 show that children of women who are White British and have an indication of poor PMH, have increased odds of risk to developmental delays in the communication (OR=2.41, CI=1.03 to 5.63, p=0.04), personal social (OR=2.40, CI=1.14 to 5.05, p=0.02), and fine motor (OR=2.51, CI=1.27 to 4.97, p=0.01) domains of the ASQ-3 30 months compared to children of women who are White British but do not have an indication of poor PMH. However, the lower bounds of the confidence intervals associated with these estimates are close to 1.00, meaning these identified associations may be insignificant in larger sample sizes.

Table 47 and Table 48 also show that children of women who are of Pakistani heritage and have an indication of poor PMH, have increased odds of risk to developmental delays in the communication (OR=3.23, CI=1.18 to 8.90, p=0.02) domain of the ASQ-3 30 months, compared to children of women who are of Pakistani heritage but do not have an indication of poor PMH. Confidence intervals associated with these estimates are wide, suggesting a high level of uncertainty around these estimates.

The p-values associated with the Hosmer-Lemeshow goodness-of-fit statistic suggested there is little evidence that the models in Table 47 and Table 48 were not a good fit.

Table 49, Table 50, Table 51, Table 52, and Table 53 stratify the regression models by child sex. Estimates are not presented for domains where there is no variation in the outcome with respect to the PMH indicator.

Table 49 Logistic regression analysis estimating the odds ratio for the risk of developmental delay in the problem solving and communication domains of the ASQ-3 27 months stratified by child sex

ASQ domain	Problem solving				Communication			
	Female children (N=183)		Male children (N=191)		Female children (N=190)		Male children (N=187)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH								
Indication of PMH	1.17 (0.37 to 3.70)	0.80	0.97 (0.38 to 2.44)	0.95	1.27 (0.37 to 4.28)	0.70	0.85 (0.37 to 1.96)	0.70
IMD decile 2019								
2	1.92 (0.59 to 6.23)	0.28	0.48 (0.16 to 1.43)	0.19	0.69 (0.18 to 2.75)	0.60	0.74 (0.30 to 1.85)	0.52
3	0.58 (0.14 to 2.37)	0.45	0.43 (0.16 to 1.21)	0.11	0.29 (0.06 to 1.41)	0.12	0.39 (0.15 to 1.03)	0.06
4	1.68 (0.30 to 9.54)	0.56	0.89 (0.25 to 3.10)	0.85	1.39 (0.25 to 7.75)	0.71	0.85 (0.27 to 2.73)	0.79
5	2.18 (0.19 to 24.97)	0.53	1.00	-	1.49 (0.13 to 16.66)	0.75	1.00	-
6	1.00	-	0.44 (0.05 to 4.07)	0.47	0.72 (0.07 to 7.53)	0.78	0.30 (0.03 to 2.77)	0.29
7	1.00	-	1.00	-	1.00	-	1.00	-
8	1.00	-	0.44 (0.05 to 4.13)	0.47	1.00	-	0.90 (0.15 to 5.40)	0.90
9	1.00	-	0.96 (0.08 to 10.79)	0.97	1.00	-	1.00	-
10	-	-	-	-	-	-	-	-
Maternal age								
<24 years	0.96 (0.31 to 2.95)	0.95	0.43 (0.13 to 1.39)	0.16	0.61 (0.19 to 1.90)	0.39	0.56 (0.20 to 1.56)	0.27
>35 years	1.15 (0.35 to 3.74)	0.82	0.59 (0.23 to 1.56)	0.29	0.64 (0.18 to 2.29)	0.50	1.21 (0.53 to 2.77)	0.65

ASQ domain	Problem solving				Communication			
	Female children (N=183)		Male children (N=191)		Female children (N=190)		Male children (N=187)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
Gestational age								
<37 weeks	0.71 (0.10 to 5.12)	0.73	1.70 (0.49 to 5.90)	0.40	0.41 (0.04 to 4.73)	0.48	0.69 (0.19 to 2.55)	0.59
Mode of delivery								
Elective caesarean section	3.90 (1.18 to 12.90)	0.03*	0.86 (0.31 to 2.37)	0.77	2.79 (0.84 to 9.30)	0.10	0.70 (0.28 to 1.76)	0.45
Emergency caesarean section	1.65 (0.40 to 6.81)	0.49	0.86 (0.28 to 2.65)	0.79	0.93 (0.17 to 5.15)	0.93	0.70 (0.24 to 2.00)	0.50
Other	3.31 (0.93 to 11.77)	0.06	0.62 (0.18 to 2.14)	0.45	2.80 (0.81 to 9.70)	0.10	0.51 (0.16 to 1.57)	0.24
NICU								
NICU admission	4.10 (1.20 to 14.07)	0.03*	1.08 (0.34 to 3.47)	0.89	1.94 (0.48 to 7.90)	0.36	1.27 (0.43 to 3.81)	0.67

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI=confidence interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 50 Logistic regression analysis estimating the odds ratio for the risk of developmental delay in the personal social and fine motor domains of the ASQ-3 27 months stratified by child sex

ASQ domain	Personal-social				Fine motor			
	Female children (N=174)		Male children (N=193)		Female children (N=185)		Male children (N=185)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH								
Indication of PMH	1.21 (0.30 to 4.87)	0.79	1.24 (0.51 to 2.30)	0.64	0.37 (0.14 to 0.94)	0.04*	0.82 (0.37 to 1.81)	0.62
IMD decile 2019								
2	1.25 (034 to 4.68)	0.74	0.91 (0.35 to 2.36)	0.84	0.67 (0.19 to 2.33)	0.53	1.03 (0.43 to 2.44)	0.95
3	0.41 (0.0.8 to 2.04)	0.27	0.89 (0.36 to 2.20)	0.80	1.39 (0.52 to 3.73)	0.51	0.58 (0.25 to 1.36)	0.21
4	1.00	-	1.26 (0.39 to 4.10)	0.70	0.87 (0.16 to 4.54)	0.86	0.76 (0.24 to 2.43)	0.64
5	1.00	-	1.40 (0.23 to 8.46)	0.71	1.00	-	1.00	-
6	1.15 (0.11 to 12.31)	0.90	0.53 (0.06 to 4.93)	0.58	0.95 (0.10 to 9.11)	0.97	1.00	-
7	1.00	-	1.00	-	1.00	-	1.00	-
8	1.00	-	5.91 (0.94 to 36.94)	0.06	1.00	-	0.75 (0.12 to 4.52)	0.76
9	1.00	-	1.00	-	1.00	-	0.46 (0.04 to 5.01)	0.52
10	-	-	-	-	-	-	-	-
Maternal age								
<24 years	0.48 (0.12 to 1.90)	0.30	0.75 (0.28 to 1.97)	0.56	0.69 (0.24 to 1.95)	0.49	0.56 (0.22 to 1.43)	0.23

ASQ domain	Personal-social				Fine motor			
	Female children (N=174)		Male children (N=193)		Female children (N=185)		Male children (N=185)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
>35 years	0.86 (0.22 to 3.40)	0.84	0.78 (0.33 to 1.87)	0.58	1.94 (0.69 to 5.47)	0.21	0.75 (0.33 to 1.71)	0.49
Gestational age								
<37 weeks	0.57 (0.05 to 6.73)	0.66	1.47 (0.45 to 4.72)	0.53	1.48 (0.23 to 9.40)	0.68	1.04 (0.32 to 3.33)	0.95
Mode of delivery								
Elective caesarean section	4.01 (1.09 to 14.71)	0.04*	1.30 (0.51 to 3.29)	0.58	2.07 (0.68 to 6.36)	0.20	0.50 (0.19 to 1.27)	0.14
Emergency caesarean section	1.28 (0.22 to 7.47)	0.79	1.46 (0.57 to 3.76)	0.44	0.86 (0.20 to 3.64)	0.84	0.75 (0.28 to 1.98)	0.56
Other	2.09 (0.46 to 9.45)	0.34	0.65 (0.20 to 2.14)	0.48	2.21 (0.66 to 7.38)	0.20	0.52 (0.18 to 1.53)	0.24
NICU								
NICU admission	2.30 (0.54 to 9.74)	0.26	1.21 (0.43 to 3.39)	0.72	0.54 (0.12 to 2.39)	0.41	1.31 (0.47 to 3.67)	0.61

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI= Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 49 and Table 50 show no statistically significant associations between an indication of poor PMH and the problem solving (OR=0.97, CI=0.38 to 2.44, p=0.95), communication (OR=0.85, CI=0.37 to 1.96, p=0.70), personal social (OR=1.24, CI=0.51 to 2.30, p=0.64), and fine motor (OR=0.82, CI=0.37 to 1.81, p=0.62) domains of the ASQ-3 27 months for male children. There were also no statistically significant associations between an indication of poor PMH and the problem solving (OR=1.17, CI=0.37 to 3.70, p=0.80), communication (OR=1.27, CI=0.37 to 4.28, p=0.70) and personal social (OR=1.21, CI=0.30 to 4.87, p=0.64) domains of the ASQ-3 24 months for female children.

Table 50 shows that an indication of poor PMH is associated with reduced odds of risk to developmental delays in the fine motor domain of the ASQ-3 27 months for female (OR=0.37, CI=0.14 to 0.94, p=0.04) and male children. Although, this association is only statistically significant for female children.

Table 51 Logistic regression analysis estimating the odds ratio for the risk of developmental delay in the problem solving and gross motor domains of the ASQ-3 30 months stratified by child sex

ASQ domain	Problem solving				Gross motor			
	Female children (N=246)		Male children (N=275)		Female children (N=246)		Male children (N=272)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH								
Indication of PMH	2.90 (1.13 to 7.45)	0.03*	1.70 (0.86 to 3.37)	0.12	2.23 (0.85 to 5.84)	0.10	1.43 (0.66 to 3.09)	0.36
IMD decile 2019								
2	1.03 (0.36 to 3.02)	0.95	1.09 (0.47 to 2.55)	0.84	0.69 (0.20 to 2.35)	0.55	1.32 (0.49 to 3.62)	0.58
3	0.59 (0.22 to 1.58)	0.29	1.19 (0.56 to 2.51)	0.65	0.71 (0.26 to 1.92)	0.50	1.62 (0.70 to 3.73)	0.26
4	0.65 (0.18 to 2.29)	0.50	0.52 (0.16 to 1.68)	0.28	0.17 (0.02 to 1.48)	0.11	0.46 (0.10 to 2.16)	0.32
5	0.22 (0.03 to 1.91)	0.17	1.95 (0.49 to 7.77)	0.34	0.72 (0.13 to 3.88)	0.70	0.64 (0.07 to 5.54)	0.69
6	0.81 (0.14 to 4.72)	0.82	0.31 (0.04 to 2.64)	0.29	0.40 (0.04 to 3.86)	0.43	0.75 (0.09 to 6.53)	0.79
7	0.29 (0.32 to 2.64)	0.27	0.29 (0.04 to 2.45)	0.26	0.80 (0.13 to 4.76)	0.81	2.19 (0.51 to 9.50)	0.29
8	1.00	-	0.21 (0.03 to 1.75)	0.15	1.00	-	0.49 (0.06 to 4.15)	0.51
9	0.87 (0.16 to 4.80)	0.88	1.48 (0.12 to 18.11)	0.76	1.09 (0.20 to 5.89)	0.92	1.00	-
10	1.00	-	1.00	-	1.00	-	1.00	-
Maternal age								
<24 years	0.58 (0.20 to 1.72)	0.38	0.86 (0.41 to 1.81)	0.69	0.72 (0.23 to 2.23)	0.57	1.24 (0.54 to 2.84)	0.61

ASQ domain	Problem solving				Gross motor			
	Female children (N=246)		Male children (N=275)		Female children (N=246)		Male children (N=272)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
>35 years	1.69 (0.75 to 3.80)	0.20	1.30 (0.61 to 2.77)	0.50	1.97 (0.83 to 4.67)	0.12	1.35 (0.57 to 3.23)	0.50
Gestational age								
<37 weeks	2.81 (0.84 to 9.44)	0.10	1.43 (0.35 to 5.89)	0.62	2.64 (0.76 to 9.13)	0.13	1.36 (0.28 to 6.49)	0.70
Mode of delivery								
Elective caesarean section	2.12 (0.84 to 5.38)	0.11	0.69 (0.26 to 1.82)	0.46	0.31 (0.07 to 1.47)	0.14	0.68 (0.21 to 2.23)	0.53
Emergency caesarean section	1.19 (0.29 to 4.93)	0.81	0.60 (0.20 to 1.79)	0.36	1.58 (0.42 to 5.89)	0.50	1.87 (0.67 to 5.23)	0.23
Other	0.81 (0.21 to 3.06)	0.76	0.61 (0.22 to 1.70)	0.34	0.58 (0.12 to 2.74)	0.49	0.80 (0.24 to 2.66)	0.72
NICU								
NICU admission	0.82 (0.29 to 2.34)	0.71	1.45 (0.58 to 3.58)	0.43	1.24 (0.41 to 3.75)	0.70	1.55 (0.56 to 4.33)	0.40

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI= Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 52 Logistic regression analysis estimating the odds ratio for the risk of developmental delay in the communication and personal social domains of the ASQ-3 30 months stratified by child sex

ASQ domain	Communication				Personal social			
	Female children (N=246)		Male children (N=275)		Female children (N=250)		Male children (N=275)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH								
Indication of PMH	4.69 (1.56 to 14.14)	0.01*	1.78 (0.93 to 3.43)	0.08	1.68 (0.83 to 3.39)	0.15	0.94 (0.53 to 1.67)	0.83
IMD decile 2019								
2	1.16 (0.39 to 3.50)	0.79	1.29 (0.57 to 2.90)	0.54	0.65 (0.24 to 1.75)	0.39	1.48 (0.66 to 3.33)	0.34
3	0.87 (0.33 to 2.29)	0.77	0.90 (0.42 to 1.89)	0.77	0.98 (0.44 to 2.17)	0.96	1.56 (0.76 to 3.20)	0.23
4	0.49 (0.12 to 2.05)	0.33	0.43 (0.13 to 1.40)	0.16	0.71 (0.23 to 2.20)	0.55	1.24 (0.48 to 3.22)	0.66
5	0.29 (0.03 to 2.50)	0.26	1.56 (0.39 to 6.21)	0.53	0.67 (0.16 to 2.71)	0.57	1.56 (0.39 to 6.26)	0.53
6	0.87 (0.15 to 4.94)	0.88	0.28 (0.03 to 2.40)	0.25	0.35 (0.04 to 3.05)	0.34	1.21 (0.28 to 5.19)	0.79
7	0.36 (0.04 to 3.40)	0.37	0.55 (0.11 to 2.79)	0.47	0.66 (0.13 to 3.44)	0.62	1.33 (0.34 to 5.14)	0.68
8	1.00	-	0.67 (0.17 to 2.66)	0.57	0.92 (0.09 to 9.83)	0.95	1.53 (0.46 to 5.14)	0.79
9	1.05 (0.18 to 6.16)	0.95	1.16 (0.10 to 14.19)	0.91	0.99 (0.23 to 4.35)	0.99	4.91 (0.40 to 59.80)	0.21
10		-	1.00	-	1.00	-	1.00	-
Maternal age								
<24 years	1.11 (0.41 to 3.06)	0.83	1.22 (0.60 to 2.45)	0.58	0.74 (0.30 to 1.82)	0.51	0.87 (0.44 to 1.72)	0.68

ASQ domain	Communication				Personal social			
	Female children (N=246)		Male children (N=275)		Female children (N=250)		Male children (N=275)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
>35 years	2.21 (0.92 to 4.89)	0.08	1.47 (0.71 to 3.05)	0.30	1.64 (0.80 to 3.37)	0.18	1.34 (0.67 to 2.68)	0.40
Gestational age								
<37 weeks	1.60 (0.45 to 5.74)	0.47	1.92 (0.49 to 7.56)	0.35	1.01 (0.30 to 3.34)	0.99	4.57 (1.11 to 18.85)	0.04*
Mode of delivery								
Elective caesarean section	0.26 (1.01 to 6.50)	0.05*	0.95 (0.39 to 2.28)	0.90	1.38 (0.57 to 3.34)	0.47	0.53 (0.21 to 1.32)	0.17
Emergency caesarean section	0.60 (0.12 to 3.06)	0.54	0.70 (0.25 to 2.00)	0.51	3.12 (1.08 to 8.97)	0.04*	0.66 (0.25 to 1.76)	0.41
Other	0.54 (0.11 to 2.55)	0.43	0.63 (0.23 to 1.68)	0.35	1.59 (0.59 to 4.25)	0.36	1.09 (0.46 to 2.57)	0.85
NICU								
NICU admission	1.57 (0.58 to 4.22)	0.73	1.97 (0.82 to 4.69)	0.13	1.13 (0.45 to 2.82)	0.79	1.03 (0.42 to 2.51)	0.95

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI= Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 53 Logistic regression analysis estimating the odds ratio for the risk of developmental delay in the fine motor of the ASQ-3 30 months stratified by child sex

ASQ domain	Fine Motor			
	Female children (N=246)		Male children (N=275)	
	OR (CI)	P value	OR (CI)	P value
PMH				
Indication of PMH	2.72 (1.24 to 6.00)	0.01*	1.13 (0.64 to 2.00)	0.67
IMD decile 2019				
2	0.99 (0.39 to 2.53)	0.99	0.87 (0.39 to 1.94)	0.73
3	0.75 (0.32 to 1.74)	0.50	1.38 (0.69 to 2.74)	0.36
4	0.74 (0.24 to 2.31)	0.61	1.37 (0.56 to 3.35)	0.50
5	0.38 (0.08 to 1.94)	0.25	2.78 (0.71 to 10.85)	0.14
6	0.31 (0.03 to 2.66)	0.28	0.87 (0.21 to 3.65)	0.84
7	1.16 (0.26 to 5.24)	0.84	0.44 (0.09 to 2.17)	0.13
8	1.00	-	0.52 (0.13 to 2.02)	0.34
9	0.30 (0.03 to 2.56)	0.27	0.97 (0.08 to 11.77)	0.98
10			1.00	-
Maternal age				
<24 years	0.89 (0.38 to 2.10)	0.79	1.46 (0.77 to 2.78)	0.24
>35 years	0.89	0.77	1.43	0.30

ASQ domain	Fine Motor			
	Female children (N=246)		Male children (N=275)	
	OR (CI)	P value	OR (CI)	P value
	(0.40 to 1.95)		(0.72 to 2.83)	
Gestational age				
<37 weeks	1.35 (0.42 to 4.29)	0.61	1.04 (0.27 to 4.11)	0.95
Mode of delivery				
Elective caesarean section	1.90 (0.81 to 4.48)	0.14	0.85 (0.37 to 1.95)	0.70
Emergency caesarean section	1.36 (0.41 to 4.48)	0.62	0.50 (0.19 to 1.31)	0.16
Other	0.90 (0.30 to 2.66)	0.84	0.66 (0.27 to 1.62)	0.37
NICU				
NICU admission	1.26 (0.51 to 3.08)	0.62	1.71 (0.73 to 4.00)	0.21

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level
OR = Odds Ratio
CI= Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Tables 51-53 show that an indication of PHM is associated with increased odds of risk to developmental delay for female children in the problem solving (OR=2.90, CI=1.13 to 7.45, p=0.03), communication (OR=4.69, CI=1.56 to 14.14, p=0.01), and fine motor (OR=2.72, CI=1.24 to 6.00, p=0.01) domains of the ASQ-3 30 months. This is also the case for male children, although, the odds are slightly lower for males than females and they are not statistically significant.

The estimates in Tables 49-53 are associated with large confidence intervals, meaning there are high levels of uncertainty, and these estimates should be interpreted with caution. In addition, the p-value associated with the Hosmer-Lemeshow goodness-of-fit statistic for the gross motor domain of the ASQ-3 30 months for male children (p=0.02) suggests that the model may not be a good fit for the data. There was limited evidence that the other models in Tables 49-53 were not a good fit for the data.

These results suggest that child sex could influence the relationship between an identification of poor PMH and risks to developmental delay measured using ASQ-3 27 and 30 months.

5.3.4.1.4 Objective Three: Differences in ASQ scores in relation to timing of PMH

As there were only 12 women who were identified as experiencing postnatal mental health, I was unable to run the planned regression analyses. This is because small sample sizes can result in biased estimates in logistic regression and there was little variation in the outcome (Newsom, 2021).

5.3.4.2 Research Question Two

Timeliness and accessibility BiB4All data

As discussed previously, primary care and health visiting data are recorded on Systm One. Data extracts are provided from Systm One to BiB for consented participants, which can then be matched to other datasets. These data extracts are sent infrequently and do not follow a schedule. This means there is uncertainty around when the next data extracts will be available. This can result in delays to accessing BiB4All data if researchers require the most up to date data extracts.

At the time when I applied for BiB4All data, the BiB data team informed me that limited ASQ data were available for BiB4All participants from the most recent data extract. It was postulated that ASQ may have not been successfully linked for these participants,

therefore, I needed to wait for updated data extracts. In addition, the BiB data team needed to generate IMD data using the routinely collected data for this analysis. This resulted in further delays in accessing these data. These issues resulted in a nine-month waiting period for the data.

Once the BiB data team received the updated data extracts, I received my requested data within one week. Hence, if researchers require data available in the current data extract or data that have already been generated, waiting times are significantly reduced.

With regards to the useability of these data, extensive data cleaning was required, as detailed in section 5.3.3.5. I also needed to request additional data, to what was provided in the initial data transfer, such as age of the participant when the code was recorded and consent date. This was to ensure the routine data related to the period relevant for the analysis and to make sense of these data. For example, the PMH data provided by the data team related to all pregnancies experienced by the BiB4All mother, including those pre-dating their BiB4All consent date. This means that not all these pregnancies resulted in a BiB4All child participant. It was important for this analysis that the identified PMH issues were experienced by the mother during a pregnancy where the child is also included in the cohort. This would allow the exploration of the association between PMH issues and child development.

Moreover, many of the values included in the datasets did not have labels. For example, mode of delivery in the maternity dataset was provided as series of numbers. I used the NHS data dictionary to label the values so that they were meaningful for analysis. Other researchers applying to use mode of delivery from the maternity dataset will also need to add labels.

Finally, I encountered issues such as multiple Mother Person IDs per Child Person ID and lack of linkage between Mother and Child participants, as discussed in section 5.3.3.3. These issues led to further delays in analysing these data as I worked with the data team to resolve these issues.

Completeness of BiB4All data

There was a small number of participants without a Systm One record (100 child participants and 24 mother participants). This is unlikely to have biased the results.

Data on child sex, ethnicity, mode of delivery, maternal age and IMD were relatively complete, where data were missing for <2% of the mother and child dyads. However, it

is unclear how accurately these data were recorded and whether they reflect the true patient characteristics and experiences.

Gestational age was missing for 11% of the mother and child dyads. As some gestational age categories used in the analysis only had a small number of participants, having these data may have had an impact on the overall estimates.

Data on whether a child was breastfed at six to eight weeks was missing for 53% of the participants. Data that were available suggested that 5% of children were breastfed and 42% were not breastfed at six to eight weeks. In quarter one of 2022/2023, UK government estimated the breastfeeding rate in Bradford at six to eight weeks to be 47% (Office for Health Improvement and Disparities, 2023). Due to the high amount of missing data, it is unclear whether BiB4All data are underestimating the true number of women breastfeeding at six to eight weeks. As the analysis planned under objective three was not undertaken, this has not impacted the results, however, this could be important for future uses of these data.

A high proportion of the ASQ-3 data were missing. At 12 months, there were less than ten children with an ASQ-3 completed within the validation window. This prevented analysis of 12-month ASQ-3 scores, which stakeholders considered important. The highest number of participants had a record of the 30-month ASQ-3, however, over 50% of child participants aged 24 months and over were without a completed ASQ-3. In addition, 1,043, 900, and 350 ASQ-3 scores for the 24-, 27-, and 30-month questionnaires respectively, were recorded outside of the validation window and excluded from the analyses. This resulted in a small sample size for the regression analysis, and left some variable categories, such as IMD, without any variation. Hence, the odds ratios estimated in section 5.3.4.1.3 should be interpreted with caution.

Another key concern with the ASQ data was that some questionnaires were completed over a number of months, for example one domain of the ASQ-3 was collected and then the other domains were collected or recorded a month later. The impact this had on the scores is unclear.

This study estimated that 70.26% of women in the cohort experienced PMH issues, using the PMH indicator. However, I was unable to access PMH data recorded in the maternity record, and I was unable to trace a woman's journey across health care services with the available data and resources, hence this estimate is not likely to reflect the true prevalence of PMH issues. Some women may have been mis-categorised, as PHM data was only recorded in their maternity record. Some women may only have discussed their mental health, and not have a PMH problem. This has implications for the analyses

presented for Objective two. Pybus *et al.,* (2023, Unpublished), explored the reporting of poor PMH in maternity, health visiting, and primary care data using data from BiB Better Start. They found that 15% of women who reported symptoms of poor PMH were identified in their maternity record. Moreover, the true quantity of the missing data is unknown as symptoms, diagnoses, and treatments of PMH noted in the free text field are not included.

5.4 Discussion

The first part of this chapter attempts to make sense of the prioritisation workshop outputs for supporting linked data research. The second part of this chapter focuses on addressing a local research priority, identified in the workshops, using linked routine data from the BiB4All study. The aim was to identify opportunities and challenges of using these data to inform local decision-making.

Consultation with stakeholders was key to conducting the research in this chapter, where stakeholders were involved in designing the research question and advised on what could be reasonably asked of the data during the analysis.

5.4.1 Defining the research question

Whilst the eligibility criterion detailed in section 5.2.1 was applied rigorously, it was not possible within the timeframe of this research to check what research had already been carried out for each of the shortlisted priorities. Nonetheless, the purpose of this exercise was to prioritise a research question to understand whether linked routine data could be used to address a local research priority.

Out of the twenty-nine research priorities that were identified across the two workshops, only three had the potential to be addressed using linked data from the BiB4All cohort. This highlights that many of the topics local stakeholders and members of the public consider important and urgent for research, could not be explored using linked routine data. The potential reasons for this are discussed in section 4.10.

As a result of coproducing this research question, it widened my network of contacts to local stakeholders in Bradford. This was beneficial when recruiting for the qualitative research study presented in Chapter 6. Those involved in these activities continued some of the conversations that we had in the session with others in the field and shared their learnings from the session about BiB4All with others.

As a consequence of developing the engagement methods set out here and in Chapter 4, a working group was developed. This was in conjunction with the Local Data Accelerator funding (see Chapter 3). This working group brought together academics and partners in the BaBi pilot sites to decide how the methods I developed may be used locally. I was a key member of this group, where I presented at the first meeting how the methods had been used in relation to my PhD and how they could be adapted for wider use. This was part of my dissemination efforts as the BaBi research teams are the target audience of this research.

5.4.2 Addressing the research question

Routine data from the BiB4All cohort study were used to explore whether there is an association between poor maternal mental health during the perinatal period and ASQ outcomes for children.

The results presented in section 5.3.4 suggest that an indication of poor PMH using this PMH indicator was associated with higher odds of risk to developmental delays for the gross motor, communication, and personal-social domains of the ASQ-3 at 24 months, the gross motor domain of the ASQ-3 at 27 months, and the problem solving and communication domains of the ASQ-3 at 30 months. Similar delays were found by Mughal *et al.,* (2019) who found that children born to mothers with persistently high anxiety symptom during pregnancy and up to three years postpartum have an increased risk of developmental delay with respect to the communication and personal-social domains.

This could suggest that if the prevalence identified by the PMH indicator is true, interventions focusing on addressing mild to moderate mental ill health during the perinatal period could be beneficial for child development, particularly for the gross motor, communication, and personal social domain of the ASQ-3 24 months, where the odds of risk of developmental delays were significantly higher (OR=4.31, OR=5.04, OR=6.84 respectively) for children whose mother was identified as experiencing poor PMH than children of women who were not identified as experiencing poor PMH. However, I would exercise caution in using these results to inform decision-making as there may be small sample bias, especially when looking at the results of ASQ-3 24 months as n<100. In addition, there was evidence to suggest that the models may be a poor fit for the data and the that the PMH did not reflect the true prevalence of PMH. Future research could explore these associations with a larger sample size to see if the

odds of risk of developmental delay are still high for children of women experiencing poor PMH compared with children of women experiencing no PMH and with more reliable PMH data. This is important as policymakers rely on good quality evidence to make decisions and this could have important policy implications with respect to PMH interventions to prevent poor child development (Department of Health and Social Care, 2022).

A potential reason why some of the estimated models may not be a good fit for the data could be that key confounding variables, such as family education, were missing from the models, as data were not available in the routine record. Further research could explore whether the model fit improves with a larger sample size and the addition of other potential confounders. The BiB Better Start Programme collects survey data on some of these confounding factors as well as data on local PMH interventions. They also have consent from their participants to access their routine data (Dickerson *et al.,* 2016). Hence, these data could then be used to re-estimate the models detailed in section 5.3.3.6 to investigate whether there are changes to the identified associations. This can inform whether collecting these data as part of the routine record would be beneficial for decision-making. Models could also be re-estimated with the removal of some potential confounding variables, such as GA and mode of delivery, as the evidence around whether they confound the relationship is uncertain (Voit *et al.,* 2022; Langham, J. *et al.,* 2023; Dachew *et al.,* 2023; Zavez *et al.,* 2021).

An interesting finding from this analysis were the statistically insignificant associations between an indication of poor PMH and fine motor development. However, it is possible that a relationship between an indication of poor PMH and the fine motor domain of the ASQ-3 at 24, 27, and 30 months may be statistically significant, with better data and a larger sample size. This is because the p-values associated with the estimates were less than or equal to 0.15. In addition, the lower bound of the confidence interval associated with poor PMH and the fine motor domain of the ASQ-30 months was equal to 1.00, which suggests there is likely to be an association.

Previous studies have shown varying effects of maternal mental health on children's motor development. Simcock *et al.,* (2016) found that prenatal maternal stress, due to exposure to a natural disaster, was correlated with better infant motor development at two months but associated with worse infant motor development at six and 16 months. DiPietro *et al.,* (2006) found that higher levels of prenatal anxiety, non-specific stress, and depressive symptoms were associated with more advanced motor development in children. This aligns with the findings for poor PMH and the fine motor development of the ASQ-3 27 months, where an identification of poor PMH using the PMH indicator was

associated with more optimal development, although, this was not statistically significant. In the systematic review by Slomian *et al.,* (2019), they also identified studies with both significant and non-significant associations between maternal depressive symptoms and infant motor development.

There are a number of reasons why PMH may affect motor coordination. Research has suggested that mothers experiencing poor mental health can be less responsive to their infants. This can lead to an insecure attachment between the mother and the infant, which can result in less encouragement for infants to explore their environment and delayed motor development (Papadopoulos et al., 2022; Nasreen et al., 2013). The effect of poor maternal mental health on motor development is shown to be moderated by infant temperament, where infants with high negative emotionality are more susceptible to the effect of maternal depression on motor development (Sacchi et al., 2018). In addition, research has suggested that high levels of maternal perceived stress in pregnancy, which is often accompanied by elevated stress hormone levels such as cortisol, affects infant brain development, which can impact on infant motor development. Cao et al., 2014 suggests that *in utero* exposure to high levels of prenatal maternal stress may be linked to cerebella dysfunction in humans, which negatively affects motor functioning.

In contrast, DiPietro *et al*., (2006) suggests that poor maternal mental health may provide a protective effect on the child's motor development, but the specific mechanism by which this occurs is unknown. They propose two hypotheses that could explain the acceleratory effects of prenatal maternal distress on development. Firstly, the human postnatal brain requires sufficient psychological stress to promote optimal synaptic structures. Secondly, more emotionally *"charged"* women may present a more labile environment in pregnancy, that includes more frequent changes to the sounds emitted from the mother's cardiovascular and gastrointestinal systems. The foetus then responds to these sensory changes caused by maternal stress. It is suggested that these additional levels of stimulation, which coincide with hormonal surges from the pregnant woman, may provide classical conditioning, which can then stimulate neural development.

Early motor functioning has shown to be related to a number of other important developmental outcomes such as linguistic, cognitive and later motor skills, as well as later psychopathology such as autism spectrum disorder (Simcock et al., 2014). Therefore, understanding the local evidence on the effects of PMH on motor development is important for decision-makers in Bradford and ensuring families are receiving the right support.

When stratifying by the child's sex, an indication of poor PMH was associated with lower odds of risk to developmental delays for the fine motor domain of the ASQ-3 27 months for female children. An indication of poor PMH was associated with higher odds of risk to developmental delay for the problem solving, communication, and fine motor domains of the ASQ-3 30 months for female children. This may suggest that the relationship between an indication of poor PMH and ASQ-3 scores is stronger for female children. If this is the case, then careful monitoring of child development may be recommended for female children of mothers with an indication of poor PMH. However, as there was a high level of uncertainty around the estimates, I would recommend caution in using these findings to inform policy decisions.

With larger sample sizes, it is also possible that these relationships hold for male children. The literature has shown differential effects of child sex on developmental outcomes. Mughal *et al.,* (2019) identified that being a male child was associated with risk of delays on all domains of the ASQ-3 except problem solving, while Kapci *et al.,* (2010) observed no statistically significant effect of child sex on developmental scores.

Sex differences might, in part, be explained by differences in maternal cortisol levels. Research has shown that maternal cortisol levels increase during the course of a pregnancy in healthy women, but that the level of cortisol varies depending on the sex of the foetus (Cao et al., 2014). Cortisol levels are higher in women carrying a female foetus from around 30 weeks gestation until near the end of pregnancy, at which point cortisol levels are similar regardless of foetal sex. Increases in maternal cortisol can result in exponential increases in infant cortisol, meaning that increases in cortisol levels late in pregnancy may have a greater impact on cerebellar development in female foetuses (Cao et al., 2014; Barrett et al., 2014). Hence, depending on when a woman experiences poor PMH, and potentially increased cortisol levels, this could explain why there are differences in development between male and female infants. However, some studies have reported that males are more vulnerable than females (Kinney et al., 2008).

Research has also shown that many women who report higher psychological stress during pregnancy continue to experience higher levels or stress after childbirth, with differential approaches to responsiveness by gender of infant (Walder et al., 2014). These differential approaches impact on the child's development. Further research is needed to understand what the impacts are for women and infants of poor maternal mental health and whether there are significant sex differences.

When stratifying by ethnicity, I found that an indication of poor PMH was associated with an increased odds of risk to developmental delays for four domains (problem solving, communication, personal-social, and fine motor) of the ASQ-3 for White British mothers.

I also found that an indication of poor PMH was associated with an increased odds of risk to developmental delay for two domains (communication and gross motor) of the ASQ-3 for Pakistani mothers. Thus, the associations between an indication of poor PMH on ASQ scores differed between ethnic groups, but an indication of poor PMH using the PMH indicator was shown to be an important predictor of child development. This suggests that interventions supporting women with mild to moderate PMH could be beneficial for child development. Small sample sizes could explain why no significant associations were found in the domains of the ASQ-3 27 months when stratifying by ethnicity. The confidence intervals associated with the estimates in these stratified models were large, suggesting a high level of uncertainty. Policymakers should be less confident in using these findings to support decision-making.

This research identified a high prevalence of poor PMH, with 70% of women having indication of mild to moderate PMH in their electronic health record using the PMH indicator. Therefore, if the associations identified in this research between an identification of PMH and child development are true, then interventions focusing on poor PMH should be prioritised. However, it is important to explore these associations with more complete and reliable data. It is possible that the PMH indicator may not be appropriate for identifying whether a woman experienced PMH, as the prevalence identified using this indicator (70%) is much higher than what is estimated by the NHS (20%) (NHS, 2022). This reflects the complexities of using routine data, that are collected with the purpose of informing clinical care, for a research purpose. Hence, I would be cautious in using these findings to inform decision-making. Further research is planned to explore how local services in Bradford are using PMH codes to record information and whether PMH issues can be captured in a more meaningful way for clinical practice and research. This planned research will explore alternative ways to capture PMH using routine data and whether this provides a more accurate representation of prevalence. This understanding will allow for better use of routine data to identify local needs, greater confidence in routine data research findings and using these to support decision making. It may also improve the chances of high quality, joined up clinical care for mother and baby. Therefore, the BiB4All cohort has provided an opportunity to explore routine data and inform how these data can be improved, so that it better meets the needs of the local community.

This research found that a lower percentage of women were identified as experiencing poor PMH for White British ethnicity than minoritised ethnic groups. This research also found that the percentage of women identified as experiencing PMH was highest among women from any other ethnic group, however, due to the availability of ASQ, I was unable to explore the associations between PMH and child development for this population.

As discussed, a woman's ethnicity and SES impacts on the identification of poor PMH, meaning that routine data is unlikely to give a true estimate of the prevalence of PMH for women from these groups and the analysis in Table 40 should be interpreted with caution (Prady *et al.,* 2015). Moreover, existing research has shown that age has an effect on whether women have a recording of their poor mental health. For example, Williams et al., (2016) found case-finding and detection was reduced for younger women. Hence, the results presented in section 5.3.4.1.2 may not be accurate in representing the prevalence of poor PMH in the BiB4All cohort.

As the number of women who were identified as experiencing only postnatal PMH was small (n=12), the results of this analysis likely represent the children of women with ongoing mental health problems or who were identified as experiencing poor mental health in the prenatal period. Therefore, policy interventions could focus on women with persisting or poor prenatal mental health as this analysis implied an association between women identified as experiencing poor PMH during these times and increased odds of risk to developmental delays. However, due to the limitations of routine data, these findings could also reflect when a woman discussed her mental health and may not accurately represent a woman's true experiences. Future research could explore the impact of postnatal mental health issues on ASQ scores with better quality and more reliable data, as this could have important policy implications.

In many of the regression analyses, there were small sample sizes for IMD deciles four to ten, meaning they were omitted from the models, or there was no variation within that category. This could suggest that the BiB4All cohort is not suitable for exploring relationships within those populations. As BiB4All was shown to be representative of eligible pregnant population in Bradford, then this may not be an issue if using these data to inform decision-making in Bradford. However, if the goal is to use these data to inform decision-making on a wider scale, then it may be useful to use linked routine data from a less deprived area.

I expected the ASQ-3 24 to 30-month data to be more complete than the ASQ-3 12-month data as health visitors routinely collect ASQ data during their 24 to 30 months check (Department of Health, 2022). However, I explored the 12-month data as this was important to local stakeholders and found there was not enough data to perform statistical modelling. Therefore, it was not feasible to use the linked routine data currently available from the BiB4All cohort to explore the associations between poor PMH ASQ-3 scores at 12 months. Future research should focus on understanding why there were high amounts of missing ASQ-3 data for the BiB4All participants. This is important given

that this is a priority locally. I hypothesise that the Covid-19 pandemic, which prevented many face-to-face appointments throughout 2020 to 2021, could explain some of these missing data, as opportunities to collect these data were reduced. The Covid-19 pandemic may also explain why some scores were recorded outside of the validation window, as appointments may have been delayed. In addition, it is likely that some children who were born in Bradford do not live in Bradford, meaning their ASQ data are not able to be linked to the BiB4All cohort. This could be explored further when data from the BaBi meta-cohort are available, as this may allow researchers to trace participants who live in different areas.

Overall, a key limitation of the BiB4All study for addressing research priorities, at the time this research was completed, was the small sample size. The small sample size resulted in high levels of uncertainty around the findings. This makes it less likely that decision-makers will use this evidence in their decision-making. To strengthen the findings, a larger sample size is needed to allow for more variation in the outcome. The high amount of missing ASQ-3 data was the main concern, and it is likely that these data are not missing at random. Understanding why data are missing should be a key research priority if these data are to inform local decision-making. Once issues surrounding missing data are addressed, these models could be re-estimated.

Another key limitation was the reliability of the PMH indicator. It is unclear whether the indicator can be used to estimate the prevalence of PMH, and additional research is needed to understand how researchers and local services can work together to develop a system that captures PMH in a meaningful way.

As the focus of this research is on the usefulness of routine data for research, I do not theorise why I identified these relationships between an identification of mild to moderate PMH and ASQ scores. This research identifies potential links between an identification of mild to moderate PMH and child development, where further research could explore this in more detail and with more robust data, in order to inform policy decisions.

Clear meta-data, including how the dataset was captured, processed, and linked, the uncertainties around the data, and how the data are defined, is important for being able to use a dataset for research (Christen and Schnell, 2023). When applying for BiB4All data, there was no clear meta-data for maternity data. This meant I had to make assumptions about these data and work closely with the BiB data team. In addition, the available NHS data dictionary provides basic descriptions of these data but does not explain how they are used in practice. Thus, I would recommend that clear meta-data are provided for all routine data, to support researchers to make decisions about which

data to request. Although, to create such meta-data, it will likely take significant resource due to the complexity of these data and the differences in how the clinical codes are used between professionals and institutions. A strength of the BaBi network is that local stakeholders are involved in the development of the cohort. This means that researchers can learn about how data are recorded, which in this case is midwifery data, and then influence those who enter data to improve quality, if this is an issue.

Currently, researchers are unsure whether routine data accurately reflect the patients' characteristics and experiences (Deeny and Steventon, 2015). Exploring the uncertainties around how routine data are collected in clinical practice could be the subject of future research, as this would better enable researchers to use these data. Specifically, understanding how PMH issues are coded by local services in Bradford is an important area for future research, given the importance of PMH to local decision-makers and communities. In this research, I addressed uncertainties around how data are used by regularly consulting clinicians, where possible. This is not always feasible as NHS staff are under increasing amounts of pressure and do not have time to engage with research. As a result, clinical input in research projects cannot always be relied upon.

Delays in accessing and using routine data for analysis can have impacts on how useful the research outputs are policy and practice. It took nine months from applying to use BiB4All data to receiving these data. Therefore, it is possible that this research question is no longer considered a priority locally, which then impacts the usefulness these findings for decision-making. This suggests that work needs to be done to allow data to be used in real time, for more pressing and impactful application to policy. However, as my data request was an external request, this experience may not be replicated for researchers working internally on behalf of policy.

5.4.3 Strengths and limitations

The consultation activities outlined in section 5.2.2 provided the flexibility to accommodate busy health professionals. The tasks during the session were chosen over other consensus-based approaches as they provided structured format, allowing one question to be chosen and for stakeholders to discuss each question in relation to a criterion. Reflecting on the consultation exercise, inviting additional public contributors may have been beneficial as there was only one public contributor compared with eight contributors from clinical and commissioning backgrounds.

By using routine data from the BiB4All study, I identified some potential challenges for using these data for research and worked alongside the BiB data team to resolve these challenges. For example, I identified that mother participants were not linked to their child in the dataset, and I worked with the data team create a process of linking them together. This will benefit researchers who plan to use data on both the mother and the child. The data team also generated IMD data for this research project, which can be made available for other researchers to use. This may save time for the data team in processing new data requests and reduce the wait time for researchers to receive their data. However, there are still some challenges that need to be overcome if these data are to be used to inform decision-making.

To make best use of BiB4All data within the scope of this project, I made some pragmatic decisions. I have been transparent about the decisions I have made, allowing this to be replicated by other researchers. For example, I chose to focus on first BiB4All pregnancies as this allowed the datasets to be easily linked. Future research could explore including all BiB4All pregnancies to see if this impacted the identified associations.

To derive variables for this research I combined information within and across datasets. This was necessary as information on a condition such as mental health is recorded in multiple health records, and by just using one dataset, you would only get a partial picture of the issue. Despite the benefits of using the composite measure of PMH, there are also a number of issues.

The code list used generate the PMH indicator is extensive but may not be exhaustive as it is challenging to draw the line between severe and mild to moderate mental health. It is also inclusive of many generic mental health codes, which may result in some women being misclassified. In addition, women experiencing severe PMH are more likely to have an indication of poor PMH compared to women with mild symptoms (Willan *et al.,* 2022). This means the analysis may reflect women with more severe PMH or it could include women with very mild symptoms as it is unclear how clinicians use PMH codes locally. When consulting with clinical experts in this project, it was important to them that we were confident that these women were experiencing mild to moderate PMH, so that decision-makers could be confident in supporting an intervention for these women. Additional research is needed to see if these relationships are maintained with more robust measures of PMH.

The PMH indicator also only indicates women who have been identified by the health service as having poor PMH. Research has shown that people do not often disclose their poor mental health to a health professional for reasons such as stigma around mental ill

health or lack of access to services (Khan, 2015; Insan *et al.,* 2022). In addition, the provision of health care services was modified during the Covid-19 pandemic and the effects of this on health data collection are unknown. It is likely that some women experiencing PMH problems were not identified due to the modification in services. Moreover, I was unable to access PMH in the maternity record. Consequently, the estimated prevalence of mild to moderate PMH in the cohort is not likely to reflect the true number of women experiencing poor PMH. This creates challenges as the exposure variable is not accurately captured. As a result, some poor child outcomes might not be associated with poor PMH, when this is the case, or these data may have shown associations when this is not the case. This has implications on how useful these data are to decision-makers. Despite these limitations, the PMH indicator was the only accessible way of using routine linked data to identify women experiencing poor PMH at the time this research was completed.

Within the scope of this research, it was not possible to determine the true quantity of missing data. For example, if people have received care outside of Bradford, this information will be missing from the BiB4All dataset. Thus, for variables such as PMH, a woman may have been misclassified as not experiencing PMH, as this was not detected in the BiB4All dataset, when they may have been diagnosed or treated outside of Bradford. Therefore, these data are likely not missing at random. This issue could be resolved by the BaBi meta-cohort, if women are able to be traced through services across a larger geographical area.

Due to data availability, I was unable to stratify the models by minority ethnic groups or by PMH timing. As research has shown that ethnicity can have profound effects on mental health, future research could explore the associations between poor PMH and ASQ scores in minority ethnic populations (Prady *et al*., 2021). Understanding the differences by ethnic group is important for targeting policy interventions. Lack of data on ethnic minority populations could also be a limitation of the BiB4All cohort for research into these populations. This could suggest that the estimates derived in this analysis are biased towards the ethnic majority.

To derive the *'breastfed'* variable used for the descriptive statistics, I used all codes that could be relevant based on their description. However, as I was unable to consult a health visitor when conducting this research, it was unclear how these codes were used. For example, is the code *'breast fed at 6 weeks'* used routinely or do clinical staff use the generic *'breastfed'* code. I also assumed that the code *'breastfed'* meant that the child was breastfed at the time the code was recorded. Alternatively, health professionals may use this code to record if breastfeeding was discussed. This further illustrates the

importance of clear meta-data. Understanding the use of these codes in clinical practice can help determine the codes requested and how these are used to derive the variables. This would increase confidence in the use of these variables in statistical analysis.

Another limitation of this analysis relates to the measure of child development (ASQ-3). As the ASQ-3 is completed by parents, rather than an assessment by health professionals, there is potential reporting bias. For example, depressed mothers may not recognise their child's abilities, which could influence the association explored in this research (Ibanez, *et al.,* 2015; Briggs-Gowan *et al.,* 1996). In addition, completion of the questionnaire can range from the parent completing the questionnaire without assistance, to the parent completing the questionnaire whilst being guided by a nursery nurse or health practitioner, as well as the parent completing the questionnaire whilst the health professional asks the parent to demonstrate these skills with the child (Kendall *et al.,* 2014). The variation in the way the questionnaire is completed might have implications for the ASQ scores and it is not possible to tell how the questionnaire was completed using routine data. This measure was articulated by local stakeholders as important evidence and was, therefore, used for this analysis.

This research is focused on a sample of the population of Bradford, meaning the findings may not be transferrable to other areas of the country. Research question one has the potential to be replicated in other local areas part of the BaBi Network, once data are available. However, as BaBi sites have different electronic patient record systems, further research is needed to explore if the data issues identified in this thesis are apparent in those systems.

The limitations discussed in the section, have implications for the usefulness of this evidence and data for decision-makers.

5.4.4 Recommendations for researchers

Based on the findings from this research, I have four key recommendations for researchers accessing and using linked routine data.

Firstly, if researchers require access to timely data, I recommend utilising data from the existing data extracts, rather than waiting for the most up to date extracts. This is because data extract timelines are uncertain, and this could introduce significant delays to the research project.

As this is the first project that has utilised these routine data from the BiB4All cohort, extensive data cleaning was required. Therefore, I recommend that researchers using

these data should plan sufficient time to clean these data. As more researchers use these data, the less data cleaning will be required as learning can be shared between researchers and the BiB data team.

Finally, I would recommend that researchers work alongside health and care professionals when analysing these data. This can facilitate a greater understanding about how clinical codes are used in practice, allowing researchers to make an informed decision about how they can be best used for research. As a result, this could increase confidence in what the findings mean for policy and practice. However, health and care professionals are likely to be time-constrained, which makes it challenging for researchers to involve them in research. A recommendation would be to include a clinician with relevant expertise as part of the project team. This would ensure that clinician input is considered throughout and encourages non-tokenistic involvement.

5.5 Chapter Summary

This chapter aimed to understand whether a local research priority could be addressed using linked routine data and to identify the opportunities and challenges associated with this. This is important for understanding whether these data can inform decision-making as part of the BaBi LHI model.

To do this, I utilised the workshop outputs from Chapter 4 and further consulted local stakeholders to define the research question addressed in this chapter. I found that many of the local research priorities were not suitable for research with linked routine data.

I then used routine data from the BiB4All cohort to investigate the chosen research priority. I found high volumes of missing data for key variables, which resulted in small sample sizes for the regression analyses. This meant there was a high degree of uncertainty around the estimates and caution should be exercised in using these results to inform decision-making. Hence, if routine data are to be used as a local health intelligence tool for child and maternal health, we need to understand why these data are missing and how we can support these data to be collected and linked. This should be a priority for further research. There was also uncertainty around the accuracy of the PMH indicator for identifying women experiencing poor PMH. Despite this, there was some evidence to suggest links between an identification of poor mild to moderate maternal PMH and their child's ASQ-3 scores, which could be used to inform future research.

Data linked as part of the other BaBi cohorts originate from different data systems and may involve different challenges. Further research could explore whether these challenges occur when using routine data from other BaBi cohorts.

The research detailed in this chapter does not aim to deter researchers from using routine data, rather, it aims to highlight potential challenges so that researchers can plan and account for these in their work, as well as work towards improving these data for future use. The next chapter explores decision-makers perspectives towards research produced using linked routine data, where the research conducted in this chapter is used as a vignette.

Chapter 6: How local early years decision-makers can be engaged and supported to make use of linked data research: A qualitative study

6.1 Introduction

The mapping review presented in Chapter 2 identified a small number of documents that described strategies for promoting the use of linked data by decision-makers. The focus of these strategies was on engaging stakeholders throughout the research process and the dissemination of research finding. However, it was unclear whether these strategies successfully increased the utilisation of linked data research by decision-makers. In addition, the mapping review identified a gap in the knowledge related to the barriers and facilitators for decision-makers to make use of linked data research. There is a wealth of research exploring the barriers and facilitators affecting the use of research by decision-makers, however, there has been little focus on the use of linked data research (Oliver *et al.,* 2014). Linked data research can face additional challenges to other scientific research such as data quality issues (as described in chapters 2 and 5), which could influence how useful these data are to decision-makers.

This chapter presents the methods and results of a qualitative study which aimed to explore how local decision-makers can be engaged in linked data research and supported to use the research outputs to inform decisions around child and maternal health. This research describes the experiences of early years decision-makers in four local areas (Bradford, Leeds, Wakefield, and Doncaster) harnessing the power of linked routine data as part of the BaBi Network. This corresponds to the evidence and decision-making stages of the BaBi LHI Model (see Figure 20). The research evidence produced in Chapter 5, as part of the earlier stages of the LHI model, informs the case study presented to participants in this qualitative research.

Figure 20 BaBi Local Health Intelligence Model*

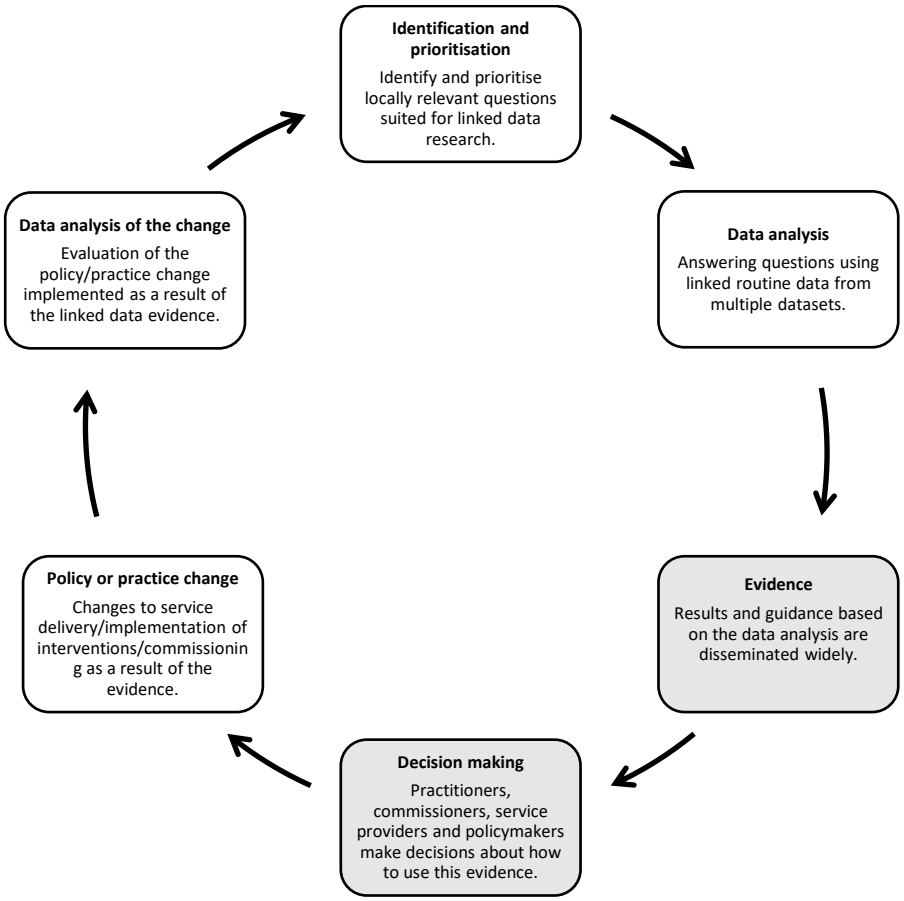


**Figure adapted from (Bryant and Bridges, 2021, Unpublished)*

Cairney and Kwiatkowski (2017) argue that for effective engagement between researchers and policymakers, actors need to understand how evidence is processed by policymakers and the environment in which they operate. As such recommendations from this research can be used to support BaBi teams to engage in activities that promote the use of linked data research by local decision-makers.

The literature surrounding how research evidence is used by decision-makers is contentious and widely studied. I have summarised this literature in section 1.5. When conducting this research, I acknowledged that decision-makers likely use multiple sources of evidence when making decisions and that linked data research has the potential to be one of these sources of information.

I use the term *'linked data research'* throughout this chapter to refer to the findings from analysis of linked routine data.

6.2 Research aim

The aim of this research was to understand how we can engage with and support local decision-makers working in perinatal and early life health to make use of linked data research as a health intelligence tool for child and maternal health.

6.3 Research objectives

The research objectives were to better understand the perspectives of perinatal and early years decision-makers towards:

1) Using the findings from linked routine data as evidence to inform the provision of services and/or development of policy around child and maternal health.
2) How they can be engaged in linked data research in order to make best use of it.
3) The engagement method developed as part of the wider project to identify and prioritise linked data research questions to be addressed with BaBi data.
4) The barriers and facilitators to using linked routine data as evidence in decision-making related to child and maternal health.

6.4 Methods

This section reports on the research process for this study, which took place between January 2022 and July 2023. It covers the study design, the critical realist epistemology and ontological foundations of this research, the sampling and recruitment strategy, data collection techniques and the method of data analysis. It also presents my reflections on conducting this research. This section is largely based on a protocol published on Figshare on the 18th July 2022 (Henderson, *et al.,* 2022a).

Feedback was sought from the BiB4All central management team on the design of this planned research as they were potential participants and users of this research. I presented my plans at the monthly BiB4All central team meeting, which I regularly attended.

6.4.1 Study design

To address the research objectives, perinatal and early years decision-makers from Bradford, and the three initial BaBi pilot sites: Leeds, Doncaster, and Wakefield, were invited to participate in an online semi-structured interview. The focus was on these local areas as their BaBi studies were furthest along in their development when planning this research and they had applied the engagement method discussed in Chapter 4.

Semi-structured qualitative interviews were appropriate for this research as they allowed for a detailed inquiry on the use of linked data for early years decision-making. They facilitated pre-specified topics to be explored whilst permitting the exploration of ideas that arose spontaneously. In addition, Oliver, *et al.,* (2014) identified semi-structured interviews as the most common method of data collection for investigating the use of evidence in policymaking in the research-policy gap literature. Hence, semi-structured interviews provided both a structured and flexible approach to understanding the participants perspectives.

Focus groups were considered for this research, as often decision-making is a collective experience and participants could benefit from recalling shared decision-making experiences (Local Government Association, 2023). It would also allow us to observe the dynamics of how decisions are made. However, focus groups provide less opportunity for detailed accounts of why individuals hold particular views (Lewis and McNaughton Nicholls, 2014), in this case towards linked data, which was important for addressing the research objectives. Participants may also feel less comfortable sharing their opinions in front of other senior decision-makers. Further to this, it would have been challenging to arrange a time suitable for a group of time-constrained health professionals or commissioners to take part, making recruitment more difficult.

An alternative study design could have been ethnographic research, observing decision-making in context. Observing how actual decisions are made rather than relying on accounts of decision-makers, could arguably create a more realistic picture of the decision-making context. However, it is unlikely that local decision-makers would engage in an ethnographic study due to the sensitive nature of their role. In addition, the findings from Chapter 2 suggested that linked data are not being utilised, and ethnographic research would likely reiterate this, rather than generating understanding around why this is the case. This makes ethnographic research inappropriate for the task. Whereas semi-structured interviews allow for probing questions to be asked about why these data may not be informing decision-making.

6.4.2 Epistemological and ontological position

Ontology, or beliefs about the nature of reality, and epistemology, the beliefs about knowledge and how its acquired, are important to consider when conducting qualitative research (Al-Saadi, 2014). A critical realist perspective is adopted throughout this project and more details of this approach can be found section 3.6 (Bhaskar, 1975).

Existing theories about how to support decision-makers to use research prompted this exploration into how to support the use of linked data research. From a critical realist perspective, existing theories may not accurately reflect reality and some theories may be more correct that others (Fletcher, 2017). According to Bhaskar (1979), we must *"avoid any commitment to the content of specific theories and recognise the conditional nature of all its results"* (pg.6). Thus, I treated existing theories about how to support decision-makers to use research as just initial theories. This allowed the exploration detailed in this chapter to either support, elaborate, or deny these theories to help build a new and more accurate explanation of how researchers can support decision-makers to use linked data research. In the discussion section of this chapter, I will outline the direct findings and also use researcher-led inference to explore how these results fit within the theories of knowledge transfer discussed in Chapter 1.

In this research, I acknowledge that perspectives towards linked data research likely differ between decision-makers across the local areas and between decision-makers and researchers. By accepting this, it allows us to understand why linked data may not be utilised in decision-making, despite the anticipated benefits for early years decision-making. Importance is placed on participants' own interpretations of linked data research and subsequent decision-making, where these interpretations are located within a particular context. This research reflects on how a person's background may have influenced their views.

6.4.3 Research timeline

January to February 2022: Application for ethical approval was successful.

January to April 2022: I applied for HRA approval as this research involves participants who work for NHS organisations. The HRA responded that approval was not needed, and the application was withdrawn.

April to June 2022: I spoke with key partners within the BaBi Network to coordinate my recruitment strategy.

July to October 2022: Participant recruitment and online interviews took place.

July 2022 to February 2023: Transcription and analysis of collected data.

July 2023: Participants had the opportunity to provide comments on a draft of the final report.

July 2023: Final report circulated.

6.4.4 Ethical approval

Ethical approval for this study was granted by the Department of Health Sciences Research Governance Committee at the University of York on 4th February 2022 (reference number *HSRGC/2022/488/A*). HRA Approval was sought but was not required for this research. The conduct of this research was monitored by my supervisory team at our monthly meetings, where it continued to be an item on the agenda. Details of the ethical considerations, in addition to those discussed in this chapter, can be found in Appendix D.

6.4.5 Sampling and recruitment

This research aimed to recruit individuals with the following characteristics:

(a) Individuals who have experience in making decisions that impact the provision of maternity, perinatal, or early years services, or who make decisions about the use of data from the BaBi cohort studies. Example roles include Director of Midwifery, Health Visitor Service Lead, maternity, children and young people's service commissioners, local authority service directors, Public Health consultants and Principal Investigator of a BaBi steering group.

(b) Individuals who work in one of the four local areas: Bradford, Doncaster, Leeds, or Wakefield, where data linkage studies have been established in collaboration with BiB.

(c) Individuals who can communicate in English, allowing them to understand the study information and provide informed consent and take part in the interview.

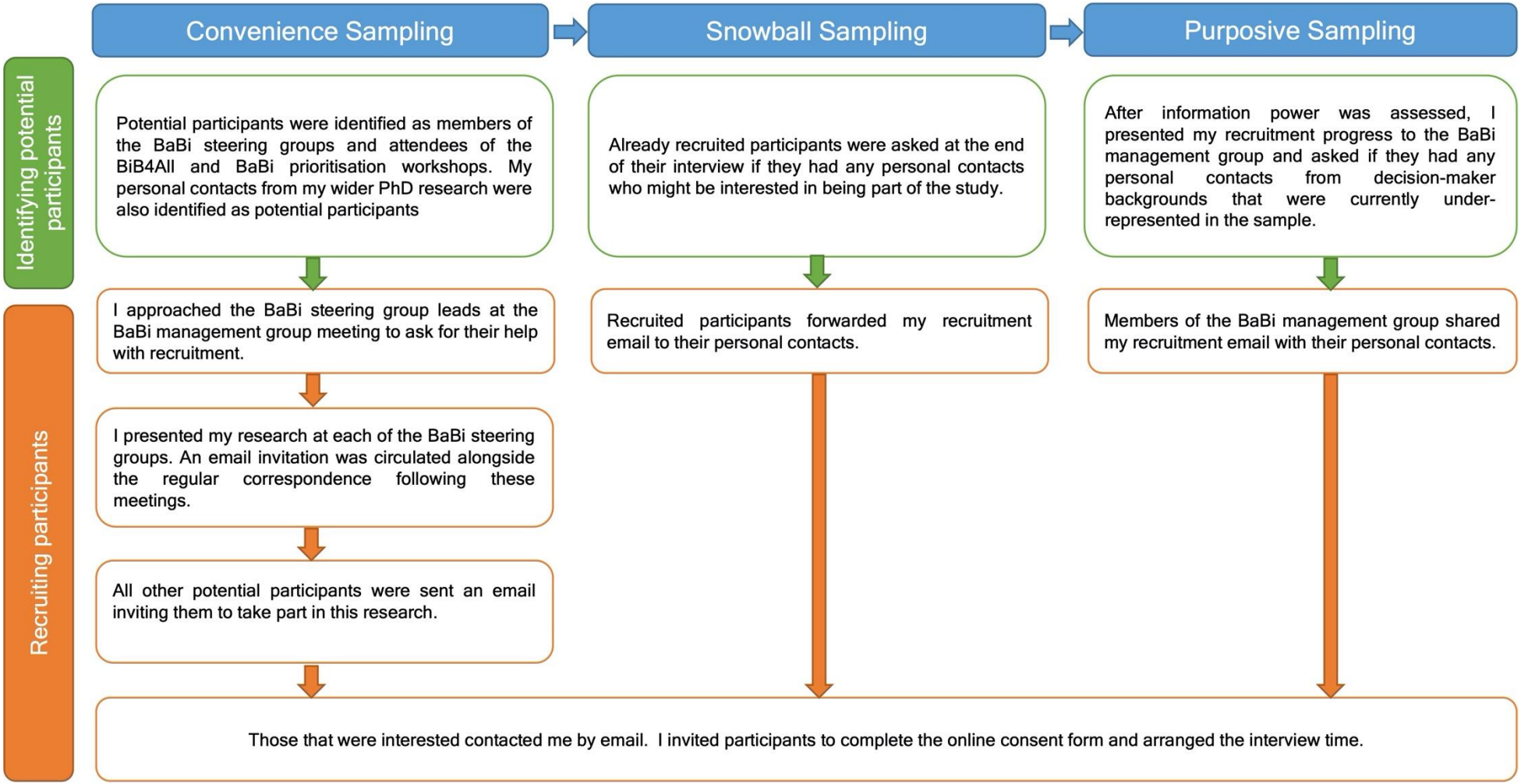
I sought to recruit between three and five individuals from each local area, representing decision-makers from a range of relevant backgrounds. This included those in broad decision-making roles relevant to early years as well as those with specialist remits, as each have the potential to make use of the BaBi data. For example, GPs make decisions that impact their local practice, which have implications for their patients. Decision-

makers within the local authority make commissioning decisions that impact on the service provision within their local area. In addition, experience of linked data research was not essential.

My aim was not to cover the whole range of phenomena, but to represent patterns in the data that were relevant to the study aim. Hence, I intended to recruit at least one representative from midwifery, health visiting, primary care, and commissioning backgrounds across the recruitment areas. Applying this pragmatic approach helped to achieve diversity across the relevant characteristics, without the need to recruit every type of decision-maker from each local area. This allowed me to explore the perspectives of decision-makers based outside of Bradford with those based in Bradford. This is significant as BiB data are established, well-known, and trusted by its local community whereas other local areas developing their studies from the start have not developed this reputation. This approach also maximised the transferability of the research findings as, from a critical realist perspective, the findings are grounded in a specific time and context. This is because it covers a range of decision-makers' experiences.

To identify and recruit potential participants for this research, I utilised multiple approaches. Figure 21 summarises the process.

Figure 21 Summary of the Sampling and Recruitment Process



In the first instance, a convenience sampling method was employed. As part of my PhD work, I built collaborative relationships with members of the BaBi steering groups, in Bradford, Doncaster, Leeds, and Wakefield through attending their monthly meetings and supporting them to conduct prioritisation workshops (see Chapter 4). Members of these steering groups were identified as potential participants of this research, as they include senior early years' decision-makers from both clinical and commissioning backgrounds. Individuals involved in leading these steering groups were also identified as potential participants and as they make decisions regarding the use of the data from the BaBi projects.

In May 2022, I approached the BaBi steering group leads at the BaBi management group meeting to ask for their help with recruitment. They were already aware of my research, so I took this as an opportunity to remind them of the research aims and reassure them that appropriate ethical permissions had been obtained. In July and August 2022, I presented my research at each of the BaBi steering groups, giving members the opportunity to ask any questions. An email invitation and Participant Information Sheet were circulated alongside the regular correspondence following these meetings. The email invitation contained my contact details, so that those who wished to take part could contact me for more information. The Participant Information Sheet and email invitation are provided in Appendix D.

Many attendees of the involvement workshops described in Chapter 4, who left their contact details to be involved further in this research, were identified as potential participants. In addition, whilst working on the research questions outlined in Chapter 5, colleagues at BiB introduced me to several clinicians and commissioners specialising in PMH, who were interested in supporting this work. Through this, I was able to identify further potential participants. These potential participants were also sent an email invitation for this research.

A snowball sampling approach was also used where already recruited participants were asked at the end of their interview if they had any personal contacts who might be interested in being part of this research. Members of the BaBi steering were also invited to share this invitation beyond the group to their personal contacts. This was necessary to ensure a range of decision-makers had the opportunity to participate and that the sample was not limited to only those involved in the BaBi Network. Thus, recruited participants could share my email invitation for this research with their personal contacts, if they would like to.

Finally, in August 2022 I presented my recruitment progress to the BaBi management group, which brings together key partners in each of the local BaBi sites. I asked if they had any personal contacts from decision-maker backgrounds that were currently under-represented in the sample and if they could share my email invitation with them. Therefore, a more purposive approach was used.

Those who expressed an interest in taking part were provided with a Participant Information Sheet for their records, a supplementary information sheet detailing the context behind the study and offered the opportunity to ask any questions. Participants were not required to read the supplementary information sheet prior to the interview as they were reminded of this information during the interview. The supplementary information sheet is provided in Appendix D.

Once the participant had confirmed they would still like to take part, informed consent was obtained. Participants were sent a link to an online consent (e-consent) form to fill out using the survey platform, Qualtrics. Qualtrics was appropriate as it allowed data to be stored securely on my University of York password protected account, thus complying with GDPR guidelines and the Data Protection Act (2018). Qualtrics also has the facility to add compulsory *'yes'* or *'no'* options to the consent questions, a box for their name and date, and a digital signature. The HRA and Medicines and Healthcare products Regulatory Agency (MHRA) joint statement on Seeking Consent by Electronic Methods was followed in the design of this consent process (Health Research Authority and Medicines and Healthcare Products Regulation Agency, 2018). Participants were also able to ask questions at any point during the process via email. The participant consent form can be found in Appendix D.

Participants were then contacted to arrange an interview time. Calendar invites were sent to participants for their interview time, which also contained the link and password to the online meeting and a guide on how to use the online platform. Participants were also advised that they could complete the interview by telephone if there were any issues with the online platform.

6.4.5.1 Information power

In quantitative research, it is common to conduct power calculations which help researchers determine the smallest sample size needed for an experimental study to detect a statistically significant effect (Kemal, 2020). Power calculations are not appropriate for determining sample sizes for qualitative research; thus, an alternative

approach is necessary. The concept of *'information power'* proposed by Malterud *et al.,* (2015) was used to determine when sufficient information had been gathered to address the aims of this study. This guided when to stop the sampling and recruitment process.

Information power suggests that the more relevant information the sample holds, the lower the number of participants that are required. Malterud *et al.,* (2015) present a model that incorporates several elements relevant to determining sufficient information power. These elements include (a) aim of the study, (b) sample specificity, (c) use of established theory, (d) quality of the discussion (e), analysis strategy. Their method implies that the final sample size should be continuously evaluated during the research process.

As I sought to recruit senior decision-makers within the NHS and local authority, there are only a small number of individuals who occupy these roles. This limited the number of potential participants for this research.

The concept of data saturation, which is defined as the point at which no new themes or codes are yielded from the data, is popular for justifying sample size in qualitative research (Braun and Clarke, 2021). The term was initially devised by Glaser and Strauss (1999) as part of the grounded theory approach to qualitative research, whereby each new observation was compared with previous analysis to identify similarities and differences. However, researchers using other analytical approaches have used data saturation without including an explanation of how this is understood outside of the grounded theory context, and why it is relevant. Multiple reviews indicate that there is a low level of transparency demonstrated by qualitative researchers regarding how data saturation had been assessed, and the concept is often poorly specified (Carlsen and Glenton, 2011; Mason, 2010; Malterud *et al.,* 2015). It is also argued that there will always be new insights as long as new data are collected, hence it is challenging to determine the point at which to stop sampling using a data saturation approach (Low, 2019). Thus, Malterud *et al.,* (2015) offered the information power concept as a way of ensuring decisions regarding design, sampling, and analysis are robust and defensible. Braun and Clarke (2021) are in support of this approach for sample size justification in qualitative data analysis.

A limitation of information power is that it still requires some prior knowledge or expertise about the approximate number of participants needed in advance (Malterud *et al.,* 2015). In addition, the items for determining information power are not always independent and they interact, which can complicate how we assess each item.

To determine whether the sample size was adequate for this research, I assessed the information power of the sample after the first round of recruitment. The seven interviews that had been conducted up to this point had high relevance to the research question and initial analytical ideas had emerged, however, I had not yet achieved the diversity in decision-makers' perspectives that I aimed for. This led to a more purposive sampling approach, where I contacted key partners at BiB and within the BaBi Network to identify individuals who occupied decision-maker roles missing from the current sample.

I conducted a further eight interviews before reassessing the information power of the data. At this stage, I was satisfied that I had captured a range of decision-maker perspectives from each local area. The sample included individuals that occupied senior decision-maker roles in early years services, meaning their experience was highly specific to the study objectives. There was variation in the types of decisions made by participants and the area of early life health which these decisions affected. This included areas such as maternal and infant mental health, parent-infant relationships, early years development, maternity services, and others, which were relevant to the study objectives.

6.4.6 Data collection

Interviews took place online or via telephone between July-October 2022. During the Covid-19 pandemic, there was widespread adoption of online platforms for business and personal communication. In light of this, the videoconferencing platform, Zoom (https://zoom.us), was used to host the interviews. The University of York have a contractual agreement with Zoom for secure and compliant data processing and storage, which covers live meetings and the storage of recordings and chat transcriptions on their cloud platform (University of York, 2023). Zoom is also viewed as a practical tool for qualitative data collection due to its ease of use, cost effectiveness, and secure data management features (Pocock, *et al.,* 2021).

Multiple research studies support online methods for conducting interviews. Abrams *et al.*, (2015) found that technology offers similar data richness in online audio-visual format to face to face. Archibald *et al.*, (2019) interviewed nurses about their experiences of using Zoom and most described their interview experience as highly satisfactory and rated Zoom above alternative interviewing mediums, including face-to-face, telephone, and other videoconferencing platforms.

However, videoconferencing is not without its challenges. Evidence suggests that working for an extended period of time online can require more concentration, resulting in online fatigue (NCCPE, 2020 and Pocock, *et al.,* 2021). Therefore, the interview time was limited to 60 minutes to minimise participant and researcher burden. There is the potential for distractions in online research as the researcher has less control and awareness of the participant's environment. Hence participants were asked to give the interview their full attention, where possible. Participants could take part with their camera on or off, depending on how they felt most comfortable. There was also a risk that participants may be overheard if they were taking part in the interview at their workplace. To mitigate this, participants were encouraged to find a quiet and private space to take part in the interview. The topics covered in the interview were not considered to be sensitive and, if overheard, should not negatively impact on the participants. There was the possibility of technical issues such as internet connection issues or inability to use the platform, therefore, a *'How to use Zoom'* guide was provided to participants in advance of the interview. Those who were unable to connect were asked if they would like to take part over the phone.

Finally, there are concerns that online methods can lead to selection bias, where participants who lack confidence in using online platforms, do not have access to sufficient technology, or an environment that allows for ethical research, are alienated. This was not a concern in this research project as potential participants of this research have access to the internet and equipment as part of their professional roles.

As the use of online methods for qualitative research are still developing, there has been little in the way of evaluation. As part of the study design, participants had the opportunity to provide feedback on the online meeting experience after the session. This assisted with the remaining interviews in this study and learning for future studies.

Interviews lasted between 20 and 60 minutes and with the permission of the participant, was recorded using Zoom. This enabled the discussion to be transcribed. A Dictaphone was also used to record the session, as a backup in the event the Zoom recording failed.

At the start of the interview, I introduced myself and my links to the project. I explained the nature and purpose of the interview and how the research would be disseminated. Participants were reminded they could ask questions at any time throughout the interview. If the e-consent form was not completed prior to the interview or if there were any concerns with the consent form, verbal consent was also recorded at the start of the interview. The verbal consent form is provided in Appendix D.

During the interview, a topic guide was followed to ensure the key issues were explored with the participants. Topic guides facilitate consistency in the data collection whilst allowing the flexibility to pursue salient details with each participant (Richie, *et al.,* 2014). The topic guide was informed by the gaps in the knowledge identified in the mapping review, the needs of the project, and consultation with partners within the BaBi Network. Topics addressed in the interviews aligned with the research objectives and a detailed topic guide can be found in the Appendix D. Interview topic guides and the Zoom platform were piloted with a senior decision-maker within the BaBi Network, where the pilot interview is also included in the dataset. The aim of the pilot interview was to ensure the generated data are relevant to the study objectives, to practise facilitator skills, and to check that the technology worked as it should.

As part of the interview, participants were presented with an example of a research question that was prioritised with local stakeholders and was addressed using linked routine data. This example question is the one outlined in Chapter 5:

If women experience mild to moderate mental ill health during the perinatal period, is this associated with their child's ASQ outcomes at ages 12 and 24 months.

They were asked if and how they could utilise this research in their decision-making and probes were used to understand why. Using an example case study can add richness to data collection by moving from general discussion, into a greater level of specificity on how they could use a specific piece of evidence (Richie, *et al.,* 2014). As this involved asking participants to think hypothetically about what they would do with this research rather than observing actual decision-making, this had implications on the data collected. For example, the relationship between how a person believes they would act in a situation and the reality of how they would act in that situation is indeterminate (Hughes, 1998). This will be considered in when interpreting the interview data.

As the researcher inevitably has an influence over the data collected, I kept a reflexive diary following each of the interviews detailing my overall impression of the interview, thoughts on how I may have influenced data collected and my assumptions throughout the interview. This was analysed alongside the transcripts of the interview.

6.4.7 Data analysis

Thematic analysis is a method of capturing relevant patterns of meaning in a set of qualitative data, using a rigorous but flexible approach. It utilises codes and coding to

develop themes and allows commonalities amongst multiple participants to be established (Braun and Clarke, 2022). What counts as a theme requires some judgement on behalf of the researcher, to determine what the important elements of the research question are.

There are a number of approaches to thematic analysis, and I have chosen the six-stage framework proposed by Braun and Clarke (2006), summarised in Figure 22. This was adopted as this study is exploratory and this approach enables a researcher to define each stage of the analysis in a flexible way.

Figure 22 Six-stage framework approach to thematic analysis*

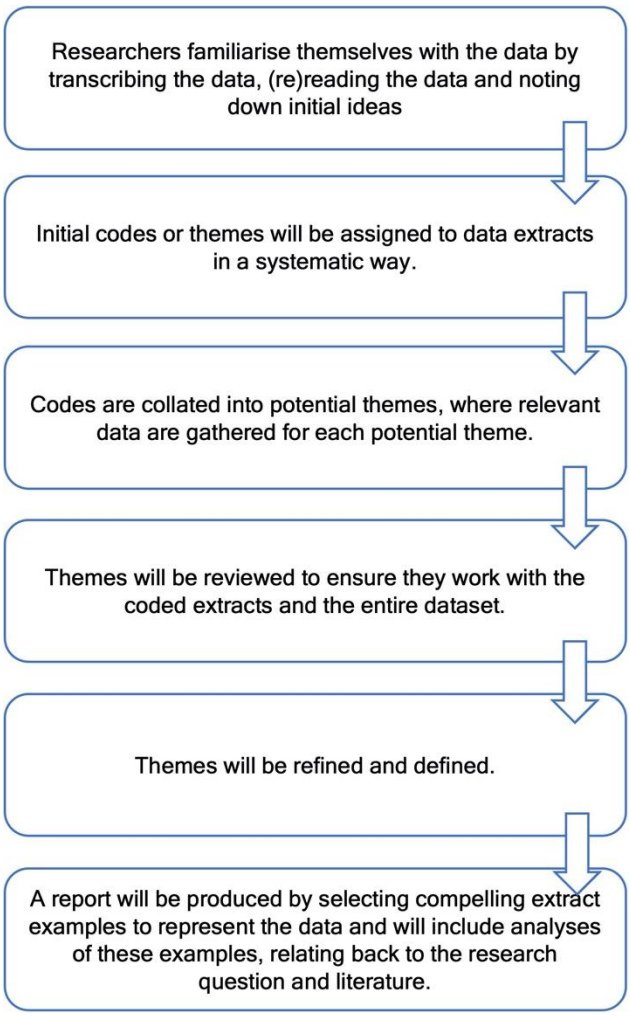


**Figure adapted from Braun and Clarke (2006).*

Thematic analysis fits within the critical realist paradigm (Bhaskar, 1975) as it can reflect reality and begin to unpick the surface of that reality as it is perceived by different individuals. This also means that the analysis of these data only makes sense when the particular conditions in which the data were generated are described (Braun and Clarke, 2006).

Themes were identified at the semantic level, where data were organised to show patterns that exist within the explicit meanings of these data, with an attempt to theorise the significance of the patterns and the implications (Braun and Clarke, 2006). This is appropriate as the purpose of this research is to produce practical recommendations for engaging decision-makers in BaBi research.

A recent editorial by Braun and Clarke (2023) provides a commentary on good practice and common problems in thematic analysis. They discuss how thematic analysis can be thought of as a family of methods, which can be applied flexibly. They also draw on their recent work on reflexive thematic approach (Braun and Clarke 2022), where researcher subjectivity is embraced, rejecting the notion that coding can ever be accurate. The editorial discussed how themes are generated and curated rather than identified or found. Thus, I considered Braun and Clarke's most recent critique of how thematic analysis has been applied, to try and avoid some of these common problems (Braun and Clarke, 2023). For example, I acknowledge that the findings of this research did not occur in a theoretical vacuum and my assumptions about the nature of reality and what constitutes as meaningful knowledge, have influenced the findings of this research. I recorded the assumptions I made when analysing these data in my reflexive diary. I have provided a detailed reflexive account of my assumptions and how I feel I have influenced the data collected and analysis of these data in section 6.4.9.

The rest of this section documents the process and my experiences of applying each stage of the six-stage framework.

6.4.7.1 Stage one: familiarisation

Qualitative data analysis begins with the researcher immersing themselves in the data to become familiar with the dataset and noticing information that might be relevant to the research question. This is known as familiarisation (Braun and Clarke, 2013).

To do this, I began by transcribing all interview recordings intelligent verbatim, which was facilitated by Otter.ai. Otter.ai uses artificial intelligence to transcribe audio recordings and stores them securely on cloud services. It also links each transcribed section to the corresponding audio recording, for ease and assessment of accuracy. Utterances were not transcribed as this research is interested in the meaning of what was said and not necessarily how it was said. Each transcript was checked for accuracy by listening to each recording alongside the text. This was the first component of familiarisation as it

allowed me to become acquainted with the information whilst noting down my initial thoughts.

Interviews were transcribed as they were completed. I felt it was important as, being an inexperienced researcher, I wanted to prioritise reflexivity and ensure integrity of the data by learning from each interview. Transcription of the interviews as they were completed also allowed me to note down the contact details for any potential participants that were suggested by the participants during the interviews, which was part of the recruitment strategy.

Once interview data had been transcribed, I read over these data collectively, as part of the familiarisation process, and noted down my thoughts and initial ideas for codes. I also read back over my reflexive diaries that described my overall impression of the interviews. As this process is observational, my initial ideas likely reflect the things most salient to me and are influenced by my positionality.

Transcripts were then pseudonymised and uploaded to a qualitative data management software (NVivo 12.1) to facilitate the remaining stages of the analysis.

6.4.7.2 Stage two: initial coding

Braun and Clarke (2013) describe coding as a process that identifies aspects of the data that relate to the research question. A code provides a label for the data, using a word or brief phrase, to capture the essence of why those data might be useful to the research question.

There are two types of coding which Braun and Clarke (2013) call *'selective coding'* and *'complete coding'.* Selective coding involves identifying instances of the phenomena of interest and selecting those out. Whereas complete coding aims to identify anything and everything of relevance to the research question within the entire dataset. This means that the researcher will code everything in the data that is relevant to the research question and then become more selective later in the analytical process.

Following the familiarisation stage, I used the complete coding approach, where I began with the first data item and worked systematically through the dataset coding small chunks of data (ranging from one line to a few sentences) relevant to the research objectives. I coded widely and comprehensively, and each data item was worked through fully before moving onto the next. Coding short excerpts of the text ensures codes are

meaningful, whilst retaining some of the context. This is not captured in strict line-by-line coding, which involves coding only one line of text at a time (Braun and Clarke, 2013).

As there is limited evidence on the perceptions of early years decision-makers towards linked data from a consented cohort, complete coding was appropriate as it enabled the themes to be data-driven. Coding inductively can protect against losing potentially important information when transitioning from codes to themes. As the interviews covered a range of topics, large quantities of codes were initially generated, but became more manageable as I progressed through the data.

Braun and Clarke (2013, pg.211) describe coding as an *"organic and evolving process"* meaning that as the researcher begins to understand the data, existing codes may be modified. As I am a novice researcher, I decided to review the codes after the first four interviews to see if any codes overlapped and that the codes were accurately describing the data items. Initial codes were reviewed, edited, and combined to make them more succinct.

6.4.7.3 Stage three: searching for themes

Using NVivo, I reviewed the excerpts associated with each code to check the code accurately represented the meaning of the data. At this stage, some excerpts were reassigned to new or existing codes that more accurately reflected the meaning.

I found the guidance on developing themes to be rather ambiguous, as there is no clearly defined method of generating themes. Therefore, to develop candidate themes from the coded data, I organised the codes into meaningful groups and related these groups to the research objectives. I looked across the groups for broader themes and discussed these with my supervisors. Sally Bridges, read the anonymised participant transcripts and developed candidate themes, which we then discussed and resolved any disagreements. I also referred to my reflexive diary as part of this process.

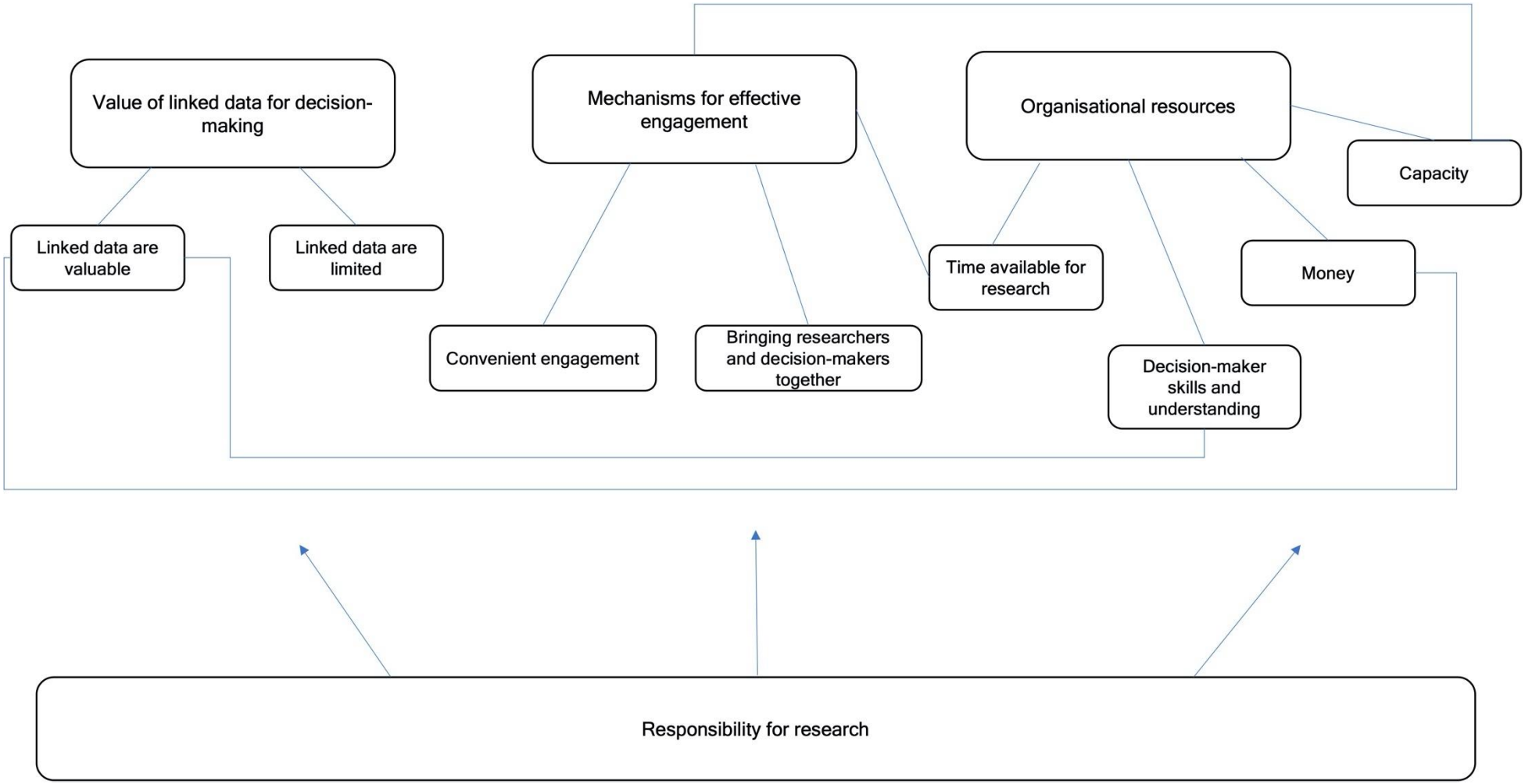
6.4.7.4 Stages four and five: reviewing and defining themes

I have integrated stages four and five as reviewing and defining the themes were part of the same process. This phase checks to see if the candidate themes fit well with the coded data, allowing you to tell a story that is faithful to the dataset. Braun and Clarke (2006) advise that themes move beyond the questions asked during the interview and represent some level of patterned response or meaning within the dataset.

As part of this process, I attempted to write up an initial set of candidate themes and realised they did not fully describe the meaning of the data. For example, I initially included the theme *'relationships'* but after attempting to write this up, it didn't quite fit with what the participants meant. When they were talking about relationships, participants described bringing together researchers and decision-makers to discuss the research. This allowed them to engage with the research conveniently and effectively, rather than having to read through many research papers. I revised these themes by going back to the coded data as well as re-reading all the data items. After the themes had been revised, I had a set of candidate themes that were distinctive from each other, worked together, related to the research question, and accurately represented the content of the interviews.

An initial theme map was developed to explore the relationships between the different themes, visible in Figure 23. The process of producing a theme map assisted in reviewing and defining the themes.

Figure 23 Thematic Map



6.4.8 Stage six: producing the final report

Writing up the analysis provided a further opportunity to review and define the themes. As I was writing up each theme, I was able to check to see if it accurately reflected what was discussed by participants.

As part of the consent process and at the start of each interview, I asked participants if they would like to be kept informed about the outputs of the research. If they expressed an interest, their details were collected and securely stored. They were invited to give feedback on the report before the analysis was finalised. This process is known as member checking, where the aim is to determine whether the results are credible and dependable from the point of view of the participants (Braun and Clarke, 2013). As a result of this process, no changes were made to the final analysis.

Where possible, I ensured that individual participants were not identifiable in the outputs associated with this research including reports and presentations. There are some instances where identification of a participant is possible (e.g., where only one person fulfils that professional role), this risk was considered in the application for ethical approval and made clear to the participant before they took part. Further precautions were taken to reduce the risk of identifying individuals such as including no quotations directly attributable to a participant, and generic job roles being used in the description.

6.4.9 Reflexivity

Critical realists argue that a researcher's knowledge, theories about the world and values can influence what is observed about the world (Fletcher, 2017). Thus, despite the rigorous coding process, the generation of themes is a highly subjective process and relies on the researcher's perception of the most striking aspects of the data. Researcher reflexivity is crucial in understanding how the themes were developed. This section describes my pre-existing ideas and beliefs surrounding this research topic that influenced the data collected and subsequent data analysis.

I began this research process with the perception that linked data research is a valuable tool for local decision-making. My views towards linked data stem from my background in quantitative research and the knowledge gained from completing the other studies that are part of this thesis. Whilst this knowledge was helpful in designing the study, as the researcher collecting the data, I was conscious of these views going into the fieldwork.

I conducted the interviews in close succession, after reading the Goldacre Report and attending a conference around linked data. This meant I was able to follow up on important issues and draw comparisons between discussions from the interviews and the wider field of data linkage. Hence, the data collected depended on the order in which the interviews took place.

I approached this qualitative research with an idea of how linked data research might be utilised locally, based on the theories of knowledge transfer, which I discuss in section 1.6. I also had some knowledge of the barriers and facilitators for using other types of research evidence, which is presented in section 1.5.1. However, I was keen to understand the views of these specific decision-makers and not to influence the discussion based on my prior knowledge, as this was key to addressing the research objectives. I believe my knowledge around linked data and familiarity of the issues surrounding its use, helped me to follow up on the salient points during the interviews.

Moreover, I assumed that decision-makers not being transparent about the use of linked data was an issue. I believe that without this transparency, it is difficult to determine whether there has been a return on the investment in linked data research, and it is difficult for people to learn from others about how to use it. By maintaining an awareness and being conscious of how my perception of evidence may be influencing the data collection, this helped ensure the data are robust.

It is possible that my pre-existing ideas about linked data research influenced how willing the participant was to share their true views on the subject. Some participants may have been reticent in sharing a view that might disagree with mine, particularly if the barriers to using linked data relate to problems with researchers. To mitigate this, I emphasised that I wanted to understand their views towards using linked data and not to worry about sharing an opinion that may differ from mine.

A researcher can be considered an *'insider'* if they share a characteristic, role, or experience with the participants of a study (Dwyer, 2009). As I have worked closely with the BaBi Network to design and deliver prioritisation workshops, I would consider myself an *'insider'*, which offered both opportunities and challenges. It benefited my research in terms of recruitment and increased participants willingness to contribute. This also allowed me to further develop a rapport with participants I had previously worked with, giving me the confidence ask to follow-up questions and respond spontaneously to the discussion. However, in some of the interviews, participants drew on shared *'insider'* knowledge, specifically in relation to key individuals within the BaBi Network and

acronyms. Hence, I followed up appropriately to clarify terms, names of individuals and also where I felt they had assumed my knowledge.

Furthermore, during the interviews I assumed knowledge of the engagement workshops from those who I had worked with on the process, and drew on shared experiences to understand their perspectives of what worked well and what did not. I remained critically aware of this *'insider knowledge'* and challenged my own assumptions whilst collecting, analysing, and presenting the results. I was committed to understanding the participants' views towards the method I developed and made it clear that they could not offend me with the feedback. Although, I accept that even with these assurances, participants may not have felt able to be fully critical and open in their responses to this section of the interview.

As this was the first time I had conducted interviews for my own research, the quality of the dialogue and the depth of the enquiry improved over the course of the interviews. Prior to conducting the interviews, I undertook qualitative research training held by NatCen (May-June 2021), Social Research Association (February 2021,) and the University of York (September 2020- January 2021) to develop my interviewing skills. The research conducted as part of my PhD research equipped me with knowledge around the use of evidence in decision-making and the research-policy gap as well as the unique challenges of linked data research, providing a good theoretical background for this research. This training and my previous networking as part of the wider PhD project allowed me to approach the interviews at ease and develop a rapport with the participants. This meant that in the majority of the interviews, there was clear communication between me and the participant.

During some of the early interviews, participants used the words, *"I don't know if this is what you want to hear"* or *"is this right",* suggesting that I influenced the responses they gave. In response to this, I was clear that I wanted to understand their perspectives and that there was no right or wrong answer to the questions. In future interviews, I was explicit about this when asking the questions, to help understand when a participant was not sure about the definition of linked data, or whether they were worried about disagreeing with my views. This issue did not tend to occur for participants who knew me before the interview.

In my initial pilot interview, I felt I could have used additional probes to elicit more detailed experiences. However, as someone who was new to qualitative research, I was conscious about remaining neutral and not imposing my own views on the participant.

The more interviews I conducted, the more confident I became in exploring interesting points in greater depth and drawing on comments made in previous interviews.

I also found that I tended towards prompts for how decision-makers could be engaged in the research, such as through prioritisation methods or involved in the data analysis, which could have influenced what the participant discussed. However, I felt that the prompts I used were mostly in response to previous discussions and that the impact of this is limited.

When reading over the transcripts, I noticed instances where I asked the participants quite lengthy questions. In these cases, participants may have only addressed parts of the question. I will take this learning into my future research projects to ensure that the questions I ask participants are clear and concise, thus ensuring they can talk about all relevant aspects of the topic.

6.5 Results

6.5.1 The participants

Fifteen early years decision-makers, working in clinical (n=8), commissioning (n=5) or BaBi Network (n=2) roles, across Bradford (n=4), Leeds(n=4), Doncaster(n=4) and Wakefield (n=3), took part in this research. They were from a range of sectors including local government, primary care, midwifery, obstetrics and gynaecology, health visiting, education, and local perinatal and early life health services. Participants included senior decision-makers in midwifery, public health leads for children and young people, Primary Care Board members, ICB representatives, perinatal service leads and senior members of local BaBi teams. Seven participants were involved in the organisation of or attended a prioritisation workshop, and eight participants were not involved in any of the prioritisation workshops. Fourteen interviews were held online, and one interview was held over the phone due to connectivity issues.

I contacted a number of commissioners in Bradford to take part in my research and three agreed to take part. However, upon following up on these contacts, I was unable to get a response to schedule the interviews. This is most likely due to time pressures and prioritisation given the political climate at the time.

To protect anonymity, all participants were assigned a participant ID (such as I01, I02, I03). I have provided details of the types of decisions that participants make, as context to the data analysis.

Clinical decision-makers who took part in this research provide care as well as decide how care is delivered. The types of decisions made amongst clinical decision-makers included:

- Types of training that are delivered to clinicians such as health visitors and midwives.
- Advising on the commissioning of perinatal services and how they are delivered.
- How to evaluate the services that are being delivered and responses to those evaluations.
- How to best support staff as well as families.
- How to best provide care within the limitations of what is approved by the NHS Trust.

Commissioners described making decisions around what services to provide, how services are structured and delivered, and which services to cut. They covered decisions related to the Best Start strategy, maternal and infant mental health services, breastfeeding support, 0-19 health visiting services, oral health, trauma and resilience, and the provision of parental education services (particularly for vulnerable families). They described making decisions collectively as part of a team and on behalf of other organisations. Senior commissioners often managed theme leads who look after child and maternal health as well as heads of services within the council.

Those who made decisions regarding local BaBi studies described their roles in setting up a local team, applying for grants and funding, organising steering groups which bring together local early years' decision-makers, engaging communities, hosting prioritisation workshops, and training community midwives.

Many decision makers were also responsible for writing reports e.g., a report on patterns and trends in health that informed the Clinical Commissioning Group's maternity strategy or policy guidelines and standards that inform clinical practice. Some participants made decisions on a small scale, such as decisions impacting their department or practice, and others made decisions on a larger scale that impacted the provision of services in a local area.

6.5.2 Themes

The aim of this research was to explore how local early years decision-makers could be engaged and supported to make use of linked data research. This section discusses the substantive themes that resulted from the thematic analysis. Four themes were

identified: 1) value of linked data for decision-making; 2) organisational resources; 3) mechanisms for effective engagement and 4) responsibility for research. The theme around responsibility for research underpinned many of the views identified as part of the other three themes. Themes related to feedback on the engagement workshop (objective three) are presented in Chapter 4.

6.5.2.1 Theme 1: Value of linked data for decision-making

This theme describes the value attributed by participants to linked data research and its use for decision-making. This includes participants' views towards how linked data can be used, its potential limitations for decision-making, and how the value of these data can be improved for future research. This theme demonstrates that decision-makers perceive value in linked data research but that there are still challenges to overcome when using these data to inform decision-making.

Participants perceive value in a network like BaBi that links routine data for research. They feel that linked routine data provides a "*more rounded opinion of what's happening with that family and what support they may need"* (I06). This can help *"tailor things to their needs, or at least understand who they are, and have a conversation about how they might want services to look different"* (I07). Decision-makers discussed how these data could provide *"better insights"* into the wider determinants of health by accessing data over a long time-period. One example related to breastfeeding and how "*we think that encouraging breastfeeding is going to make a difference to a child's oral health… However, the way that data is currently captured, … you could understand outcomes related to breastfeeding, but you couldn't then connect that to any wider determinants of health, so… future propensity to have good oral health. And I think that potentially some of this [BaBi data] will allow us to have better insights into that"* (I10).

This perceived value suggests that linked data have the potential to inform local decision-making as it can be used to improve the way local services are delivered. Decision-makers described how the data could provide specific insights into *"earlier support and earlier help"* for families as "*the biggest issue that we have is, we sort of get involved with these patients too late… by the time the kids get to school, and they can't read"* (I09).

Linking data across services was also described as a way of "*encouraging partnership work"* between local services, as "*when you're expecting partners [data] and not just your own data to inform you … I think it helps everyone see that whatever it is you're commissioning has got a benefit for lots of different reasons" (I15).* This can help

decision-makers "*to work more seamlessly across what had previously been silos"* (I12) and it supports "*buy-in for joint commissioning type opportunities"* (I15) to address early life health issues.

An example of the partnership work related to *"a joint commission between prevention in early health and public health"* (I15). The commissioning teams can *"see from the data, what is the issue right now"* and use it to *"spot where you're going to change it and whose priorities or strategies, that's going to be"* (I15). This suggests there is value in not just the data itself, but in how it can generate this buy-in and influence to improve the health of local families.

Moreover, participants expressed how *"it's a bit of a hindrance at the minute that we don't have that type of data"* (I07) and how *"it's always been on our mind, how data should always be linked up"* (014). As they have "*data potentially sitting in silos in different organisations",* this makes it challenging *to "understand someone's journey and to understand what works from a prevention perspective*" (I07) in the absence of linked data. This suggests they perceive value in linked data as it allows them to access the benefits described.

Participants also suggested designing a system that allows linked data to "*be embedded in our commissioning decisions… so it almost becomes routine, that you would consult this data before you make a decision"* (I10). This allows the perceived value of routine data to be realised.

When participants were given a specific example of research being conducted with linked data and asked whether this could inform policy or practice, it was clear that participants felt this research could be valuable for decision-making. As noted by participants, they considered this research as "fundamental" and "vitally important" in ensuring their local mental health service "exists" as it "*stops us having to argue that we need this money to offer this service to prevent problems arising in the future, because it would prove that there are already problems right now"* (014). It can also help commissioners "*to think about how would we want to target [specific funding] and what sort of interventions would we want to deliver"* to ensure that money is spent "*in the right way to create better long-term outcomes and reduce health inequalities"* (I10). Decision-makers from all BaBi sites and professional backgrounds described how useful the findings could be to them.

Furthermore, decision-makers suggested that BaBi research could be used as *"a starting point"* or *"a flag"* (I07) to prompt further research or discussions. BaBi research was described as "*observational in nature and not interventional in nature"* where decision-

makers can explore "*the sorts of interventions which go on typically from day to day"* to identify patterns and trends, *"and then return back and devise intervention studies… to see whether the observed interventions actually do have benefits, or otherwise in properly designed interventional trials"* (I08). It was suggested that "*although one notices interventions in an observational study, they're not randomised, and therefore, they can only be used for hypothesis generation, in terms of what interventions are going to be beneficial" (*I08). Thus, *"the data itself isn't the solution"* (I07) and further research such as randomised studies are needed in response to these data.

Despite participants describing the value of linked data for decision-making, it was recognised that "*there's lots of issues in the system in terms of how data links in together, that would be very handy to iron out"* (I14). What information is captured, how accurately routine data are recorded, and the consistency of recording information across services and datasets were issues highlighted by participants regarding linked data. These issues were discussed by participants from all local areas and professional backgrounds, and were perceived to limit the value of the data for research. Participants spoke about the data from a personal perspective, where many of them had been involved in data collection or oversee a team which collects the data that are linked as part of BaBi. This reflects their own frustrations with the poor-quality health record systems.

Linked routine data were described as *"limited"* as a result of poor templates used to code these data in electronic health systems and incorrect recording of the data by professionals.

> *"Whilst you … could get some really good data from the coded stuff, because the codes are not particularly well done, you might … only be able to get that a question was asked, rather than anything more …. there's some things that I know aren't coded, so then how will you actually be able to then follow what happens next?"* (I11).

Several participants used the example of ethnicity to illustrate the challenges of linking data across systems.

> "*Just comparing our ethnicity data with maternity's ethnicity data, we're collecting it using different, kinds of, codes. So, we end up, it's very difficult to do a comparison when you're comparing apples and pears"* (I11).

Thus, it was suggested that *"it only really makes sense as linked data if everyone's using the same language"* (I14).

These issues with using linked routine data for decision-making could be the result of decision-makers not being "*involved in designing the templates that people are coding in the first place"* (I11). Participants recommended "*getting the people in charge of data from each system in a room"* (I14) to agree on a consistent way of recording the data.

It was also suggested that educating clinical professionals at "*undergraduate level, pre reg level*" (I03) on the importance of inputting the data correctly and the value these data will have, could improve future routine data collection. This could include developing an understanding of how the electronic patient record "*means more than what it does on the day you're looking after your patient"* and the importance of getting "*your patients' data correct"* (I03). An example of how routine data could inform decisions about emergency departments was used to illustrate the importance of accurate data collection "*say [name] was in A&E [Accident and Emergency], on this day, [their] blood pressure was this and actually, we know all this stuff about [them], this helps us think about, is the ED [Emergency Department] in the right place?"* (I03).

Finally, participants expressed concerns regarding the consent process for BaBi and *"if someone doesn't consent, then you're completely taking them out of the study"* (I01). This creates challenges if *"you get like a group of women who… might not speak English very well, … they might not understand what research is… so, you might get a lot of people not consenting from those kinds of groups. But they're the kind of groups that you want in your research, because you want them to be well represented"* (I01). This can impact on the value of the research depending on *"what [decision-makers are] trying to decide"* (I01).

Hence, linked data are perceived as a valuable tool for decision-making although concerns regarding the quality of the data and the limitations of the systems that record this information need to be addressed.

6.5.2.2 Theme 2: Organisational resources

Human and organisational resources, including time, capacity, money, and skills, were discussed as factors that enable or prevent decision-makers from making use of linked data research. The extent to which resources limit decision-makers ability to use linked data research was discussed both in terms of producing linked data and acting on the findings it generates.

There is an apparent tension between resources and requirements, where resources are described as influencing decision makers' ability to respond to emerging research.

> *"obviously, we've then got to look at resources to see if we can generate a service or provide a service that meets those needs"* (I09).

Participants talked about how *the "affordability"* of the intervention dictates whether they are able to act on the research findings.

> *"But then it was left to local areas to know whether they could afford it. … But actually, if you've got your local evidence, right, affordability follows it because you're then not wasting your money, doing something a bit random building a community centre, in an area that nobody wants a community centre, because somebody thinks it's a good idea"* (I03).

An example of some research that used routine data to explore the association between preterm birth that placed the child in an early school year and school readiness, was used to illustrate how decision-makers tend to action research if it does not take up too many resources.

> "*schools and local government can look at those data and say, yeah, we believe that it's not, it's not involving a whole tranche of babies every year, that's, that's doable. Let's make, let's make that happen"* (I08).

Participants spoke about being "*time poor"* and how this influences their ability to engage with research outputs. Time was described as a *"luxury"* and *"precious"* to decision-makers and that "*there's only so much time any of us can put [into] research, whilst also understanding that it's kind of giving us so much learning that we want to participate"* (I15). Thus, time pressures faced by decision-makers is a perceived barrier for decision-makers to use linked data research in decision-making. Time available for research directly influences their ability to access, understand, and action all the research. This can be considered part of the policy-making context and was described by both clinical and commissioning decision-makers. This links to the theme around mechanisms for effective engagement as participants described easy and accessible engagement methods that allowed them to engage with research given their time pressures.

Linked routine datasets are relatively new to local decision-makers and a number of participants were not confident in their understanding. When asked to describe their understanding of linked routine data interviewee 14 responded, "*my understanding is … probably not anywhere near as good as it should be."* (I14).

Developing decision-makers' understanding of linked data and linked data research were frequently discussed by participants, where this was considered an enabler to using linked data research. Education around linked data focused on three key areas: 1) making sure the right people know the linked data exists, so that it can be factored into decision-making; 2) the people who want to use it know how to access it; 3) people know how to use it, which includes knowledge about the content and integrity, analytical skills, and how to apply the findings in practice.

> *"What enables [decision-makers] a) they know about it, b) they can access it and c) they can understand the… integrity of it, so they know it's robust and safe, and they can use it, it's not made up, in the best way and they understand how to use it"* (I03).

Participants described the need for *"clear meta-data"* and understanding about where the data are going to be hosted, for example "*it's going to ultimately be hosted in a TRE [Trusted Research Environment]"* (I10).

Educating policymakers was considered important, as it ensures "*that part of our community who we trust, to look after our money, and to spend it…, you know, our council leaders, or our national leaders have the right evidence upon which to base those decisions"* (I08).

There were varying degrees of confidence expressed by participants regarding stakeholders' ability to understand how linked data could be used. Some felt *"comfortable"* in their understanding of how linked data could be used and described having *"enough capacity and capability within the team to engage in [conversations around linked data] in a meaningful way."* (I12). Participants suggested that "*people in public health would be pretty well versed in what kind of data is being collected"* as it is *"kind of [the] basis of everything we do"* (I15).

Others expressed the need for further support to build this understanding, and how they would benefit from developing an understanding of *"the types of data that there are. Because I think you often know what's available in your own sector, you don't necessarily know what's available in other people's"* (I07).

There was interest from some decision-makers in developing skills in linked data research.

> *"I'm like, just feel like a lifelong learner, I always think that it's really good to keep learning new stuff…. I've never been a researcher, so, it's really helpful, isn't it to spend time with people who work in a different discipline to you to get some insights into how these things operate? And what the challenges are?"* (I05)

This suggests efforts to increase knowledge of linked data would be well received.

Several participants talked about using examples of how the data have been used in other local areas to build this understanding, to help facilitate decision-makers to use these data. For example, "I *know when I've been to various [decision-maker meetings] to talk about BaBi … I've used the air pollution examples that Born in Bradford have done… how useful that might be for our decision making … because then it was a concrete example of a research outcome"* (I13).

6.5.2.3 Theme 3: Mechanisms for effective engagement

This theme explores participants' views towards how they want to be engaged, in order to make best use of the research outputs. There were many elements involved in participants' views of effective engagement. These included engaging the right people, in the right context, with the right methods and outputs. An overarching factor of effective engagement was convenience.

> *"I suppose it's just finding the right people for the right conversation, at the right time is really important for us"* (I15).

Participants describe being *"bombarded with information" or "in a mound of reading"* where "*it all blurs into one"* (I02). Therefore, they *"need [research] to be quick to engage with"* to "*make use of all the intelligence coming up"* (I15). This links to the time pressures described under the organisational resources theme. It was suggested that bringing together researchers and decision-makers to discuss the research findings can overcome these challenges.

For example, participants suggested having "*an open workshop for any sort of stakeholders, staff, etc, within that area, so that [researchers] can present [the research]. And so, it gives you a chance to ask questions and get clarity around what it means. And how we're going to work together to then implement any actions" (I06).*

Alternatively, researchers "*come and give an update at something like [an early years] meeting?... and then … anyone that has questions, can ask them there and then, because it is easier when you meet a person, isn't it... as opposed to just reading something"* (I02). This also facilitates key players to work together to implement actions in response to the research.

> "*And I think you can only do that by getting the key players in a room, can't you? And allow them to be presented with the research, be able to ask questions about the research and then take time then talk about well, what, how you going to use it to make a difference? What's each stakeholder going to do"* (I06).

An example of how bringing researchers and decision-makers together can be an effective method of engagement was described by a decision-maker who was interested in a research project focusing on an intervention for babies and young children. The project had published a number of research papers and the decision-maker wanted to utilise the research, but due to time constraints were unable to process all the information. The participant described how being able to speak to the researchers directly about the research enabled them to make better use of the findings.

> *"[The researchers] have published [a number of] different research papers, one about implementation, one about feasibility, one about acceptability… But I have to read like 23 million different research papers. So, instead, what I did last week was email them to say, could we have ten minutes to just chat through exactly what all of these tell me and if you had x amount of money, what you would do? And so that's what we did…"* (I15).

They described how they wanted "*all of [the researcher's] expertise to tell me all of those discussions and conclusions" and "If [they] had 100,000 pounds to spend on [specific area of health] on children in [city] right now, what would you do"* (I15).

This implies that decision-makers want researchers to provide a convenient way for them to access the information they need to make decisions. This links to the previous theme as decision-makers describe convenient ways of being engaged because they have limited time.

The idea of hosting *"knowledge exchange type events"* was also suggested as a way to bring researchers and research users together to share research but also learning and skills around how others have utilised the data to inform decisions. This would involve having "*people who are actively involved in using the BaBi data…" and "people actually,*

kind of, sharing what they did, what they found, how they used, how it was used in policy and practice and the opportunity for local government officers and decision makers to meet the academics and sort of get to know each other a little bit" (I13). This was described as an enabler for decision-makers to use linked data research.

Participants made specific suggestions about who the appropriate decision-makers would be to present the research to, using the example research study that I presented on perinatal mental ill health and children's ASQ outcomes (see Chapter 5). For example, the research could be presented to "*members of the [perinatal mental health] group. So, those practitioners that work with families, whether it be the NHS, voluntary sector"* (I02) to discuss.

Moreover, it was considered important that the research was presented using the right method and in the right format. This is because "*all the stakeholders that you want to engage, tend to be quite busy, important people who tend to sort of be pulled in all kinds of directions and really have to kind of prioritise their time"* (I01). Many clinical professionals are *"treading water to stay afloat. And because of that, if you turn around and say right, you need to give up so much time to organise this, I think you'll get a lack of engagement"* (I09).

Hence, accessible and convenient engagement methods that are not burdensome of decision-makers' time were described by participants as a way of convincing decision-makers to get involved in the research. For example, having a quick conversation with a researcher.

> *"Oh, I've just been looking at this, it's really piqued my interest, can I discuss it with you for ten minutes? That might be a really good way of doing [engaging decision-makers in research]"* (I10)

Participants expressed that giving stakeholders "*enough notice to schedule [research engagement activities] into your diary*" (I05) was important.

> *"As long as you give [clinical staff] six-week notice, just because that's how they run their worlds."* (I15)

Moreover, "*a really specific question set for a meeting or consultation type exercise is really important, so you get the right person"* attending. This is because *"if you call it data linkage, you could end up with a [team member] who's all about the data, but not necessarily about deciding whether perinatal mental health data or language data is like*

a priority… they wouldn't be able to bring you that. So, it's the getting the question, right?" (I15).

Some participants offered to take the research to the relevant decision-makers as a way of effectively engaging decision-makers in the research outputs. It was suggested that researchers should regularly present at the relevant decision-maker meetings and ask them to identify which groups new research should be presented to.

> *"There's a [specific] committee that sits within the [organisation]…And I'd be more than happy to support where it feels helpful... so there's kind of a formal route, that kind of is like, yes, I would like it to come to these [meetings] regularly. But then there's a more like responsive route, where I'm saying, actually, if you ask for my help, I can help to connect you into the places where this information needs to go"* (I10).

The view that research should be presented to decision-makers was shared by both BaBi decision-makers and non-BaBi decision-makers across local areas.

Different modes of disseminating the research were suggested by participants, where each involve a short time commitment from the decision-maker. For example, presenting the research as a *"short video or infographic"* or as *"summary type information",* where they could "*really quickly link [the research] into some of the priorities and things [they are] working on"* (I10) or "*go back to the original research papers if we want to see what the actual kind of data was"* (I15). This is effective as it means decision-makers are presented with the main headlines and only need to explore the research in more detail that is related to their decision-making context. This was important "*because we all get busy. And as [name], who's a [data analyst role] in the council said, she feels like she turns her head and something else pops up. Which is true"* (I15). Presenting the research in a convenient and accessible format was described by decision-makers from both commissioning and clinical backgrounds and across local areas.

6.5.2.4 Underpinning theme: Responsibility for research

When asked about how they could be engaged and supported to use linked data research, participants discussed research as either part of their decision-making role or as an additional responsibility. This underpinning theme represents how the different types of decision-makers perceive their role in research and how these perceptions influence engagement in linked data research.

Those decision-makers' who view research as the responsibility of academics, struggle to find the time to engage with linked data research as they prioritise other responsibilities.

> "*Most healthcare professionals couldn't give two hoots about research. It's, it's such a sort of seen as so that's an additional extra to their daily lives"* (I08).

> "*It's the capacity, it's whether they've actually got the, the time outside, you know, given other pressures that they have"* (I13).

This contrasts with those who feel it is part of their role or feel passionately about it and make time for it, often in addition to other competing priorities. This suggests a tension between the time they have available and what is a requirement of their job. This tension is demonstrated in the following quote as the participant explains how they feel obligated to engage with research but that many decision-makers don't have the "*luxury of time*" to do so.

> "*I feel in a really fortunate position because I have loads of time…allocated for reading and catching up on evidence that's out there. And I know my clinical colleagues don't have that luxury a lot of the time, and they're really busy…But they're still under the same obligation as I am to revalidate and to demonstrate that I know what's happening locally, regionally and nationally, in terms of evidence"* (I05).

There is a strong theme that research is considered outside of decision-makers' core work, that it is *"academic",* and separate from their *"practical job"*. Participants implied that researchers are not only responsible for doing the research, but for suggesting how it can then be implemented. This is apparent in the theme: mechanisms for effective engagement, where participants describe how researchers can make their research accessible and convenient. They explain that this is because decision-makers are "*time poor"* and although they'd "*love to spend all [their] days kind of reading [research]… that's not the nature of [their] job, because [they are] not an academic"* (I10). Its suggested that *"the nature of [their] job is that it's kind of busy"* and that the role of the research team should not be undervalued "*because actually, [researchers have] got this kind of luxury of time and the luxury of not being embedded in a day to day thing, which is about decision making, or managing teams… so actually, [they] need [researchers] to do some of our thinking for [them]"* (I10). Therefore, the view that research is the responsibility of researchers and not decision-makers, underpins how decision-makers want research to

be presented to them. This again suggests a tension between time and the requirements of their job.

Multiple decision-makers described how clinical professionals may not see the value in being involved in this research. As a result, it was suggested that some clinical groups "*aren't represented as well as they should be [in research], and that's not through … the research people not inviting us… I was choosing to be engaged, because there's a lot of [clinical professional group] that if there's not any gain in it for them, they won't necessarily give up the time, because time is a precious commodity."* (I09). This implies that unless they perceive research as part of their role and not an additional extra, then they will not feel they have time for it. Hence, it was suggested that BaBi research can provide "*good springboards into discussions about … what research actually looks like"* to demonstrate how research is "*everybody's responsibility"* (I08).

6.5.3 Online method

Eight participants filled out the feedback form regarding the use of Zoom to conduct the interview. They were asked how they found conducting the interview online and all responses were positive. The most common response was *"fine".* When asked about their preferred way of being interviewed 50% chose Zoom, 37.5% chose Microsoft teams and 12.5% chose *"any method"*. When asked to explain this, one person who chose Zoom said it was *"easy to access the meeting online".* For those who chose Microsoft teams, they said that this was the main system used within their organisation. Seven participants expressed that their interview experience could not have been improved. One person said their interview experience could be improved as there was a technological challenge but that I managed this well. Based on this feedback, using a range of platforms to host online interviews could be considered and the online platform preferred by the participant could be chosen. This is appropriate as the different organisations have different online platforms approved for use.

6.6 Discussion

This section presents the key research findings, the implications of these findings, the strengths and limitations of this research, and the research impact.

6.6.1 Key findings

Participants perceived linked data as a valuable tool for local early years decision-making. Decision-makers across all local areas and professional backgrounds were supportive of the BaBi project and expressed they would be interested in using the research outputs. They discussed the benefits of linked routine data as part of BaBi, which included more information on local families; facilitating a joined-up approach to improving services; and access to early years data. Participants described wanting linked data to be available and that data could be used to explore connections between the wider determinants of health. This implies that BaBi can fill gaps in current knowledge.

It was expected that these decision-makers would attribute some value to linked data given the recent large investments in data linkage made by local government (Ministry of Housing Communities and Local Government, 2021). Participants of this research were either involved in a local BaBi steering group or were contacts of those involved in a BaBi steering group, meaning that many of the participants had heard of linked data and its potential benefits. Local BaBi teams were also working with partners across the NHS and local authorities to promote the BaBi project, which may explain why participants attributed value to linked data. The people who took part in this research were likely those who were interested in linked data. Moreover, the example research question used during the interviews was co-produced with local stakeholders, therefore, it was expected that it would be relevant locally. Understanding that local decision-makers perceive value in BaBi research is significant as this means it has the potential to be used as a local health intelligence tool for child and maternal health. This suggests value in further efforts to promote the BaBi project locally and engage decision-makers in the project, to ensure the research produced is relevant to their needs.

However, participants also expressed how electronic health record systems used to record routine data can limit the quality of the data for research and decision-making. Therefore, improving how data are recorded electronically is recommended. Participants suggested that educating health and care professionals during their professional training about the importance of accurate routine data collection for research and local decision-making could increase the value of the data, by overcoming some of the data quality issues discussed. If you educate the people at the beginning of the chain about the value of these data, then this could impact the value of these data for the people at the end of the chain, the decision-makers. This is because those that input the data are collecting data for clinical or administrative purposes rather than research purposes, meaning there may be less of an incentive to input data that are not relevant to their clinical problem. If health and care professionals understand how those data can be used for research and how this research can then be used to improve clinical practice, it may incentivise more accurate collection of those data. The perception that linked routine data are of poor

quality could be a factor that contributes to why limited evidence of linked data informing early years policy and practice was identified in Chapter 2.

Nonetheless, the benefits of educating those that input the data are limited by the data system. To improve the collection of routine data, the systems in which the data are recorded need to be developed to ensure they can accurately capture the information that is important for research.

Data quality issues associated with linked routine data were identified as barriers to research use in *Whole system data linkage accelerators: a North-South partnership to unlock public health data* project (Wright, 2022, Unpublished) and by the Organisation for Economic Co-operation and Development (OECD) (2019) in their *Health in the 21st Century: Putting Data to Work for Stronger Health Systems*. A recent unpublished study explored the views of data users, stakeholders, and community members in Bradford and Tower Hamlets (Wright, 2022, Unpublished). This research covered their views on the challenges of providing, accessing, and using linked data as part of Connected Bradford and the Whole Systems Data Project. Participants indicated that it would be useful to have descriptions of the datasets and how they can be accessed. This would help data users understand the context of how the data was collected and the limitations of the data. This was consistent with the discussions from this research.

The juxtaposition of these views towards the value of linked data research could demonstrate a lack of understanding of what linked data are. People are assessing the value of these data based on little knowledge of what is captured and how it can be used. In describing the benefits of linked data, participants commented that linking data across services could be used by practitioners to identify individuals who are accessing multiple services, and tailor those services to their needs.

> "*If you can have the data from all of those services and understand how it's impacting an individual or group of individuals, and I think it's much easier to make sure you can tailor things to the to their needs, or at least understand who they are, and have a conversation about how they might want services to look different"* (I07)

However, routine data available as part of the BaBi Network are anonymised, meaning it cannot be used to identify these individuals to talk to them about their needs. These expectations of the data may be overstating the value of linked routine data for decision-making. Thus, it is important to understand if these data are still useful to decision-makers if these benefits cannot be achieved.

Concerns over data quality could also be explained by a lack of understanding of what is captured in the routine record. Many participants were not confident in their understanding of linked routine data, which could be because linked data are relatively new to some local areas. This lack of understanding may lead to an underestimation or overestimation of the value of linked routine data. Participants suggested that developing a decision-maker's knowledge and understanding around linked data would enable more effective engagement with linked data research.

If decision-makers underestimate the value of linked data research, they may be less likely to use this to support the commissioning of a new service or seek this type of evidence. For example, many participants suggested that the linked data research could be used in conjunction with other evidence, where linked data research is used as a prompt for further research. This aligns with Weiss' (1979) theory of enlightenment regarding how research is used in policymaking and is typical in the literature for research utilisation (Dobbins *et al.*, 2007). Participants described how linked data could be used to identify whether a problem exists in their local population, that could then be explored further through trials. The possibility of conducting trials within the BaBi cohorts or using the data to evaluate trials, as has been demonstrated in the data linkage literature, was not discussed (Roblings *et al.,* 2021). Participants associated linked routine data with observational and not causal research, which could suggest that decision-makers attribute more value to controlled studies (such as randomised control trials) and do not associate these controlled studies with linked data research. Further research could explore this in more detail. If this is the case, this could limit the potential value of linked data for decision-making as there are missed opportunities to use these data to conduct trials within cohort studies (Relton, *et al.,* 2010). This further highlights the importance of educating stakeholders on the uses of linked data research.

The wider evidence-based policy literature discusses how policymakers' skills influence research utilisation (Oliver *et al.,* 2014). Participants of this research refer specifically to decision-makers' knowledge and understanding of the datasets that are available as part of BaBi and how they can be utilised. Developing these skills is likely to be a huge task, given the breadth and complexity of the datasets. Providing a description of these data is also unlikely to be accessible to decision-makers due to this complexity. Education on linked data, beyond data descriptions, is likely to have a greater impact on decision-makers ability to engage with linked data research. Further research is needed to understand what training on linked routine data should look like.

Participants suggested that using examples of how linked data have been used by others could be an effective way of building this understanding as well as hosting knowledge exchange events between researchers and relevant decision-makers. In Switzerland, eHealth Swiss has published guidance for educators on how to integrate digital health topics into education and professional training of health workers (OECD, 2019). Learning from this initiative could be used to develop training for local policymakers.

To further support decision-makers to make use of linked data research, participants discussed effective ways of engaging them in BaBi research. Key to this was convenience and a short time commitment required by decision-makers to engage with the research. This reflects the busy nature of decision-makers roles discussed by both small and large scale decision-makers.

Participants described bringing together the right people, at the right time, to discuss the outputs of the research. This is an effective use of decision-makers' time as they can ask researchers questions about the findings that are relevant to their decision-making context, without the need to engage with all the research outputs. It also provides the opportunity to discuss potential solutions that relevant decision-makers can then implement. Contact, collaboration, and relationships between researchers and decision-makers are also identified as important facilitators of *'research use'* among policymakers in the broader literature (Oliver *et al.,* 2014; Contandriopoulos *et al.* 2010; Innvær *et al.* 2002; Mitton *et al.* 2007; Nutley, *et al.* 2007; Walter, *et al.* 2005; Orton *et al.,* 2011). It is suggested that contact and collaboration between researchers and decision-makers can establish trust and enable discussions around what the important policy issues are and how research fits into that (Oliver *et al.,* 2019). Although the discussions presented in this chapter centre around engaging decision-makers in the research outputs, decision-makers also expressed they would like to be involved in the prioritisation workshops to help shape the research (see Chapter 4). This demonstrates that participants want to be involved in research over the project lifecycle.

Participants also communicated that they wanted research to be presented as a short summary, alongside the full research paper, allowing them to quickly engage with the research. This again reflects a short time commitment to the research. This is consistent with the existing literature on the use of evidence by decision-makers, where it is recommended that research be presented in shorter formats, written in plain language, freely available, and in multiple ways to address the needs of different audiences (Cairney and Oliver, 2018; Dobbins *et al.,* 2007; Lavis *et al.,* (2005); Contandriopoulos *et al.* 2010; Innvær *et al.* 2002; Mitton *et al.* 2007; Nutley, *et al.* 2007; Walter, *et al.* 2005).

The discussions imply that improved communication between researchers and decision-makers regarding research outputs can support the use of linked data by decision-makers. This echoes concepts outlined in the *'two communities'* model of knowledge translation and Boswell and Smith's (2017) *'knowledge shapes policy theory'*, as the findings suggest that understanding and communication are factors that influence research use. Boswell and Smith's (2017) *'knowledge shapes policy theory',* focuses on the perceived gap between research and policy communities, where relevant research does not achieve impact due to communication problems. As such, research may not be presented or disseminated in a way that is relevant or accessible to policymakers or the changes required cannot be implemented due to resource constraints. This could also explain why the mapping review in Chapter 2 identified limited evidence of linked data research informing early years policy and practice. The findings from this research and the wider literature suggest that better access to research can be achieved by improving the communication between the two communities. Further research is required to understand whether communicating the research outputs in the way participants describe, makes a difference to the uptake of the research in policy or service provision.

Time constraints and lack of capacity were frequently discussed as barriers for decision-makers to engage with linked data research. These factors were also identified by Oliver *et al.,* (2014) as broader barriers to research uptake by policymakers. Some participants describe how they lack the time to engage in research, whilst others recognise research as part of their responsibility. Where research is seen as additional to their core role, this creates challenges for engaging these decision-makers in linked data research. This is because these decision-makers may prioritise tasks they perceive to be part of their role over engaging in research. It also implies that some decision-makers assume that it is the responsibility of the researcher to promote their findings in a way that decision-makers can easily action. There is a gulf between raw routine data, the headlines of the research and what the solutions are based on the findings, and it is unclear where the responsibility for bridging this gap sits.

Understanding that some decision-makers feel research is not part of their role, helps us to appreciate why they may prefer convenient methods of engagement and allows us to develop methods of engagement that recognise this. As such, presenting research in a way that decision-makers can quickly and effectively engage with, could encourage greater engagement in linked data research.

In a systematic review by Orton *et al.,* (2011) regarding the use of research evidence in public health decisions, three studies explained that policymakers were not supported through training or the organisational structure to use research evidence. This could

explain why some decision-makers' do not see research as part of their responsibility. This suggests that to change the utilisation of linked data research by decision-makers, it requires a change at the systems-level to ensure research is better embedded. Further research is needed to explore if and why some decision-makers see research as additional to core responsibilities and if this is the reason why they are not willing to give up their time for research. This may aid the design of more effective communication and collaboration strategies.

As BaBi is a collaborative project at the local level, it is important that everyone is engaged and feels responsible for its use and the success of the project relies on this. The introduction of the ICS provides an opportunity to embed research within the disparate health and care system. Research is a key theme running through the ICS' strategies, where partners across the system are encouraged to work together to deliver research. The Health and Care Act 2022 sets out legal duties for the Integrated Care Boards (ICB), which includes the facilitation and promotion of research in areas relevant to the health service, and the uptake of evidence from research. NHS England will assess ICB on their performance with respect to these duties (NHS, 2023a). In addition, the NIHR has recently provided funding to establish HDRCs to embed a culture of research and evidence-based policymaking within the local authorities (National Institute for Health and Care Research, 2023). Thus, the narrative around the responsibility of research in health and care is changing and it is moving towards a system where everyone is accountable for research. Based on the discussions from this research, this change in narrative could benefit the use of linked data for research and health care decision-making, as research becomes a more embedded consideration for decision-makers.

As a result of these findings, if linked data research from the BaBi studies is going to influence local policy and practice, the research outputs need to be presented in the right format and to the right people. However, if we don't address decision-makers' assumptions around their responsibility for research and the disconnect between researchers and policymakers, then these efforts to engage decision-makers in research might not make a difference. This is because decision-makers will continue to see research as not part of their core responsibility. Thus, bridging the gap between research and policy communities is important if linked data research is to inform decision-making. This echoes the recommendations from the *'research-policy gap'* literature presented in Chapter 1.

The themes identified in this research overlap with those identified in the mapping review presented in Chapter 2. For example, Davis-Kean *et al.,* (2017) suggested that

policymakers need results more quickly to be able to react to imminent issues and they often need data from intervention studies before implementing them. Macfarlane *et al.,* (2019) also describe issues with quality around the Hospital Episode Statistics submitted by some maternity units, which could also be a barrier to using linked data research. Pitching the research to the right decision-makers was also identified by Hopf *et al.,* (2014), which is consistent with the findings of this research.

Much of the analysis discussed in this chapter likely applies to research more generally and is not limited to linked data research. As such, some of the recommendations provided could improve the use of research more generally by early years decision-makers, as well as the use of linked data research.

6.6.2 Strengths, limitations, and challenges

A diverse range of early years decision-makers across NHS and local authority organisations, in four areas that are part of the BaBi Network (Bradford, Leeds, Doncaster, and Wakefield), were interviewed as part of this research. This generated understanding about how potential users of linked data research perceive linked data and its use for early years decision-making, as well as how they want to be engaged in this type of research. This addresses an important gap in the existing knowledge base, identified in the mapping review detailed in Chapter 2. This research was timely as BaBi studies were launched in Leeds, Doncaster, and Wakefield with an organised media appearance in July 2022, when the interviews for this research began.

I interviewed both small and larger scale local decision-makers to understand if they had different needs when engaging with linked data research. I also interviewed people who make decisions about the BaBi project in their local area as I aimed to understand how they could be supported and how the current engagement method worked in their local areas. Their experiences broadly resonated with the experiences of decision-makers not as closely involved in the BaBi project.

Due to the implementation of the new ICS that took place whilst I was completing this project, decision-makers often covered larger areas. For example, Leeds, Bradford, and Wakefield are all part of the same ICS. Therefore, the sample of decision-makers needed for this research was reduced as decision-makers were working across multiple local areas. In addition, some of the research participants described having multiple decision-maker roles within their organisations and had experience making decisions in several

areas of early life health. The richness of these data resulted in a smaller overall sample needed to address the aims and objectives of this research.

The outputs of this research reflect the views of the decision-makers that took part. It was likely that those who decided to take part had an interest in linked data. Hence, the sample is likely to be biased towards those interested in using linked data. Decision-makers in local areas that were not included in this research, may have differing views towards the use of linked data. This is in line with a critical realist perspective. Therefore, there is likely to be more variability across all local decision-makers than was observed in this study. However, the themes identified in this research align with those found in the wider literature on the use of evidence by decision-makers but are more specific to linked data research. Consequently, the recommendations from this research could be useful to other local areas part of the BaBi Network.

Cairney and Oliver (2018) have been critical of the advice provided by researchers for improving research use, by suggesting it is not informed by policy studies or accounts of the relationship between evidence and policy. The research detailed in this chapter encompasses anecdotal evidence, from a decision makers' perspective, on how they want to be engaged and what has worked well for them previously, to add to this body of literature. The recommendations from this research can be used to better design the engagement strategy for teams working with linked data research. The mapping review in Chapter 2 found limited evidence of linked data being used in policymaking, thus it was necessary to approach this research in this way. When the landscape has changed, further research may lend itself more to exploration through case studies or accounts of actual policymaking.

Using the prioritised research question from Chapter 5 as a vignette during the interviews allowed participants to relate to a linked data research study that they have the potential to use to inform decision-making. This led to a more detailed discussion around the value of linked data research and how they wanted the research to be presented.

Overall, hosting the interviews online went well, however, I did encounter some challenges. A number of the attendees could not access Zoom using the NHS system or on their work laptop. This meant that participants joined the calls later than expected, which resulted in shorter interview times. One interview was hosted by phone call to overcome the challenges of Zoom. If I was to conduct this research again, I would explore Microsoft Teams as another potential online platform for participation. NHS digital has also assured the use of Microsoft Teams, and this is the chosen videoconferencing platform (NHS, 2023b).

In anonymising the interview transcripts, I faced the challenge of balancing two competing priorities: maximising the protection of participants identities and maintaining the value and integrity of the data. I was keen not to downplay the role of the decision-maker whilst also trying to keep their identity hidden. As there are relatively few people who occupy senior decision-maker roles, this makes them potentially identifiable, and participants were made aware of this risk during the consent process. As participants' job roles were highly important to understanding their experiences, it was crucial that this information was retained for analysis and by completely removing this information would undermine what they are saying about their decision-making. Thus, I have been unable to include copies of the transcripts as part of this thesis as it may reveal a participant's identity. However, researchers wishing to use the data collected as part of this study can apply to the BiB executive committee.

6.6.3 Implications and recommendations

The aim of this research was to produce practical recommendations for teams setting up studies as part of the BaBi Network on how best to engage local decision-makers in the project and in the outputs of their research.

Based on these interviews, I would recommend establishing short, regular communication between BaBi teams and local decision-makers, that facilitates discussion of research outputs but is not burdensome on both parties. Continuous interactions over time can allow relationships between decision-makers and BaBi teams to form, allowing for more effective engagement (Oliver *et al.,* 2014). For the existing BaBi sites, this could involve communicating with the local services they are currently partnered with and whose data are going to be linked.

Methods of communication could include *"knowledge exchange type events"* or an "*open workshop",* as suggested by participants. Opportunities for decision-makers to meet with researchers to ask questions on an *ad hoc* basis could also be beneficial, where the experiences of I15 have demonstrated success with this type of engagement. For these regular meetings, the agenda could be set in advance and sufficient notice provided, to allow relevant decision-makers to attend the sessions. Further research could inform the format of these interactions.

Decision-makers who took part in this research suggested that the quality of the data could impact the value of the data for decision-making. Based on the suggestions put

forward by participants, I would recommend the BaBi teams collaborate with local universities to ensure that professionals who input routine data are educated during their professional training on the importance of routine data for research and decision-making. This could encourage more accurate coding of data in the electronic records, improving the data quality for future research. I would also suggest working with companies who design the electronic record systems to ensure they are capturing the information that is relevant to clinicians, commissioners, and researchers. Further research should explore if there are any barriers to improving the collection of routine data.

Education on the uses of linked data and its potential value should also be provided to decision-makers. This research identified that some participants were not confident in their understanding of how the data could be used. By developing this understanding, data from the BaBi studies could be used more widely to inform a greater range of decisions. This would also enable decision-makers to better engage with linked data research, including the prioritisation of linked data research in the workshops described in Chapter 4. As developing this knowledge will take place over time, I would recommend that BaBi teams hosting prioritisation workshops or engaging decision-makers in the research, provide more informative descriptions of the available datasets in BaBi and how they can be accessed, so that decision-makers can begin to develop this understanding. Some decision-makers were keen to undertake training, implying that efforts to educate decision-makers would be well received. Further research is needed to understand how to best educate decision-makers that is both effective and requires a short time commitment. This is important as decision-makers who do not develop their understanding of linked data are less likely to use linked data research to inform their decision-making. This would be a missed opportunity for better informed early life health decision-making.

It was unclear from this research whether decision-makers' perceptions towards their role in research influences the time they were willing to devote to linked data research. Further research in this area is recommended, given the importance the new ICS structure has placed on embedding research across the health and care system. Changing the narrative around who is responsible for research has implications for how research is utilised within the health system. If decision-makers perceive value in being involved in research, this could improve the use of linked data research by those decision-makers. In the short term, I would recommend that BaBi teams identify those stakeholders who have a greater interest in research and maximise this potential opportunity for engagement.

Some of the recommendations presented in this section, are analogous to the *'how to influence policy and practice'* advice found by Oliver and Cairney (2019) (see section 1.5.4). This advice includes communicating research effectively with clear summaries allowing policymakers to follow-up if they have questions and the importance of building trusting, long-term relationships between researchers and policymakers.

6.7 Dissemination and impact

The findings of this project are most relevant to those developing studies that link information about mothers and babies across multiple public services, for research purposes. Therefore, the findings have been distributed to the BaBi Network. The recommendations from this research can be found in the BaBi Network Google shared drive. I plan to also present this research to members of the BaBi Network during the regular meetings. Moreover, the findings informed the development the engagement method detailed in Chapter 4.

The final report was circulated amongst the participants who expressed an interest in staying informed and to the wider community. This research has been presented at the HSR UK 2023 conference which is attended by academics, clinicians, and members of the public. I plan on publishing this in an academic journal and sharing the final report with the Early Life and Prevention (ELP) Theme of the YHARC.

As a result of taking part in this research, some participants expressed an interest in being involved in the research further. One participant asked to be invited to the next workshop in their local area. Another participant suggested including some of the BaBi data in a proposal for funding, as the research question I am addressed in Chapter 5 of this thesis is relevant to their bid. They also said that they found the conversation *"really helpful",* and that taking part had made them think more about how they might use the data. Hence, this research has raised awareness of linked data as part of BaBi and its uses, which can indirectly affect research utilisation. This is because decision-makers will be thinking more about how this research can work for them. Some participants asked questions about the project, which may have helped develop their knowledge of linked data.

6.8 Chapter summary

This chapter explored the perspectives of early years decision-makers, in four areas that are part of the BaBi Network, towards using linked data research to inform local decision-

making. Participants were asked about how they could be engaged and supported to make use of linked data research, with the aim of informing the engagement strategies within the BaBi Network. This research was important as there is limited evidence of linked data research informing early years decision-making in the UK, despite the potential benefits for health and care services. It also provided the opportunity to gather feedback on the engagement method used to generate research priorities, detailed in Chapter 4, from the perspectives of those who are the intended contributors and beneficiaries.

The rise in evidence-users, within the UK healthcare setting and the increased investment into linked data studies, makes this research timely. The use of linked routine data for research is relatively new to most local authorities and a cohort like BaBi has not previously been established. Therefore, it was important to understand how decision-makers view this type of research, if BaBi is to be used to inform local decision-making as part of the BaBi LHI Model.

Semi-structured interviews support this type of understanding to be generated by allowing detailed, individual accounts of experiences of linked data. Thematic analysis of the interview transcripts identified four themes: value of linked data for decision-making; organisational resources; mechanisms for effective engagement; and responsibility for research.

This research implies that although these local decision-makers perceive value in linked data research, there are a number of barriers that need to be overcome if these data are to be used as a local health intelligence tool for child and maternal health. This research indicated that local decision-makers in areas part of the BaBi Network want to be engaged conveniently by reading short summaries of the research and discussing the outputs with researchers. Effective engagement methods are necessary due to the time constrained nature of decision-makers' roles and potentially their views towards their responsibility for research. This implies that linked data research is more likely to be valuable to decision-makers if it is convenient and accessible. To support these local decision-makers to use linked data research, alongside other sources of evidence, it was recommended the quality of the routine data is improved by tailoring the data collection systems and educating those who collect the data on the importance for research and decision-making. Further research is needed to understand how to effectively educate decision-makers and professionals who input routine data on the potential of these data.

The outputs from this research can inform how studies linking data for the purposes of early years research can engage and support local decision-makers to make use of their

research. This allows for better informed early years decision-making and effective use of existing routine data.

Section C: Discussion

This section contains a discussion of the results.

- Section 7.1 summarises the key findings in relation to the research objectives and implications of these findings.
- Section 7.2 describes the strengths and limitations of the research conducted in this thesis.
- Section 7.3 provides practical recommendations for BaBi sites as part of the thesis aim.
- Section 7.4 makes recommendations for future research and Section 7.5 makes recommendations for policy and practice.
- Section 7.6 provides a chapter summary.

Chapter 7. Discussion

7.1 Summary of findings

Chapter 2 of this thesis presents a mapping review, which aimed to identify gaps in the literature to inform the research conducted in this thesis. The mapping review highlighted a gap in the knowledge for how linked data research is used by early years decision-makers in the UK and effective strategies that support this. It was unclear whether linked data research was being used to inform early years decision-making but was poorly documented, or if decision-makers face barriers when using this research.

Hence, the primary aim of this thesis was to understand if and how linked data can be used as a local health intelligence tool for child and maternal health, using the BaBi Network as a case study. The overarching aim was to generate a greater understanding about how both researchers and research users could be supported throughout this process.

To do this, I explored the following objectives, which correspond to four stages of the BaBi LHI model:

1. To identify research priorities around child and maternal health to be addressed using BaBi data (stage one of the BaBi LHI model)
2. To explore whether a local research priority can be addressed using linked routine data from the BiB4All study (stage two of the BaBi LHI).
3. To understand the perspectives of local decision-makers towards evidence produced using linked routine data and using it as a local health intelligence tool for child and maternal health.
4. To identify the support needs of local decision-makers to use evidence produced using linked routine data as a local health intelligence tool for child and maternal health (stages three and four of the BaBi LHI model).

This discussion revisits these objectives by summarising the main findings of the studies within my thesis. Exploring these objectives allowed a greater understanding around the opportunities and challenges for linked routine data to be generated, thus contributing to the wider knowledge on the use of linked routine data in decision-making.

7.1.1 Identifying priorities for linked data research

Chapter 4 of this thesis presented a method of engaging local stakeholders to identify research priorities that could be addressed using BaBi data. Identifying research questions for linked data research is the first stage of the BaBi LHI model and is part of the foundational phase of the FHI 360 Research Utilisation Framework. Public and stakeholder engagement during this phase is important as it allows researchers to understand the needs of local services and communities, so that the research outputs are relevant to the issues faced locally (Kim *et al.,* 2018; INVOLVE, 2012). The mapping review in Chapter 2 also identified the importance of engaging decision-makers early in the research process.

Existing priority setting methods were not suitable, hence I developed a pragmatic two-hour online prioritisation workshop method, which has now been successfully applied in five of the BaBi sites. This method was developed as part of an iterative process, where learning from each application of the method was incorporated into the workshop design. A guide on how to host the online prioritisation workshop is included in a toolkit to support new and developing BaBi sites and will continue to be developed following the completion of this PhD. Positive feedback on the workshop format was received from attendees and those who took part in the qualitative research in Chapter 6. Hence, this method can be useful to teams prioritising questions for linked data research, and there is the potential for this method to be developed to facilitate priority setting across different areas of health care.

This thesis demonstrated that it is possible to engage people from a range of backgrounds to identify local priorities for the BaBi studies to address. However, the priorities that were identified were broad topic areas and these needed further refinement into research questions, as I describe in Chapter 5. Therefore, the workshop guidance recommends that facilitators focus on gaining clarity on a smaller number of research ideas, to allow the workshop outputs to be more useful to local BaBi teams.

In addition, many of the research priorities were not suitable for research using linked routine data and were better suited to qualitative research or research that has access to bespoke questionnaire data. This could suggest that linked routine data are not capturing all outcomes that are relevant locally. This was also discussed in the qualitative research detailed in Chapter 6, where participants discussed how the limitations of the data systems restrict the usefulness of routinely collected data for research and local decision-making. Participants also discussed how decision-makers are not involved in designing the templates that are used to record routine information, which could explain why these data are not capturing the relevant outcomes. Therefore, working collaboratively to identify research priorities and examining the available data, provides

an opportunity to work with local services and data providers to ensure these important outcomes are routinely captured. This is discussed further throughout this chapter and is a key recommendation of this thesis. The research priorities not suitable for linked data research could still be addressed by the BaBi studies by using BaBi's *'consent to contact'* mechanism.

Moreover, increasing people's knowledge and awareness of linked routine data and its potential uses would likely help workshop attendees and facilitators to better identify priorities that are suitable for linked data research. It was apparent in the qualitative interviews in Chapter 6 that participants wanted to develop their understanding of linked data research, to better enable them to use these data. This is discussed further in section 7.1.4. During the workshops discussed in Chapter 4, attendees also mentioned that they would like more information about the data that are available in the linked dataset, as this would help them understand what they can reasonably ask of the data. Therefore, it could be beneficial to spend time developing an understanding of datasets linked as part of BaBi, as this would allow for more effective priority setting for linked data research. This is important if BaBi data are to drive decision-making, as we first need to identify research priorities before these data can be used. Without research priorities, researchers cannot move to the research phase of the research utilisation process. If researchers did set their own priorities, they may not reflect the priorities of local decision-makers, meaning they are less likely to inform policymaking. Exploring ways of developing a greater understanding of linked data would be a valuable subject of future research.

7.1.2 Addressing local research priorities using data from linked routine data

To the best of my knowledge, my thesis includes the first study to use linked routine data from the BiB4All cohort for research. I used these data to address a research priority that was identified by local stakeholders (Chapter 5). I aimed to lay the foundations for other researchers to access and use these data to support children and families. I also intended for this research to demonstrate the potential of using these data to support decision-making.

This thesis identified a number of challenges when using these data to address a specific local research priority. One of the main challenges related to small sample sizes. For example, ASQ-3 24-, 27-, and 30-month data were available for <50% of BiB4All children who were aged 24 months and over. This was unexpected given that ASQ scores should be routinely collected for children in the UK by health visitors as part of the Healthy Child

Programme (Department of Health, 2022). There are a number of reasons why these data may have been missing. For instance, parents may not have taken up their health visiting appointment, where these data would be recorded. This was likely to be compounded by the Covid-19 pandemic, where parents may have been concerned about attending a face-to-face appointment. There is also the possibility that children moved outside of Bradford before their ASQ was recorded, meaning these data would not be available as part of BiB4All.

Small sample sizes resulted in a large amount of uncertainty around the estimates. Good quality data are needed to support decision-making and lack of confidence in the research findings limits the usefulness of these data for decision-making. However, the results did imply a relationship between an identification of mild to moderate mental ill health during the perinatal period and children's ASQ scores, which indicates this could be an important area for future research. This suggests that the BiB4All data can provide only partial intelligence for decision-makers on the impact of maternal mental health on children's ASQ scores, until these data are available and data quality are improved.

Another challenge related to the reliability of the PMH indicator for identifying the prevalence of PMH within the cohort. This suggests that significant work is needed to understand how data are being recorded in practice and for what purpose, as this can guide how researchers work with local services to ensure routine data are captured in a way that is meaningful for research and decision-making.

Data quality issues were also highlighted by decision-makers who took part in the qualitative research detailed in Chapter 6 and by health and care professionals who were consulted throughout this project. This is in accordance with the broader literature that describes the limitations and challenges of using routinely collected data from population-based studies (Robling *et al.*, 2021; Harron *et al.,* 2017; Warren-Gash, 2017; Lugg-Widger *et al.*, 2018; Raftery *et al.,* 2005; Herzog, *et al.,* 2007; Davis *et al.*,2016). This research highlights the need for better quality routine data, especially around key outcomes such as the ASQ and PMH, if it is to be used for research and to inform local decision-making. Without good quality data, researchers are unable to rely on the research findings to make recommendations for policy and practice. Hence, if linked data are to be used as a local health intelligence tool for child and maternal health, we need to focus on improving the collection of routine data for research. As this research has highlighted a gap in the ASQ data, understanding why these data are missing could be a key priority. Researchers could:

- Explore how missingness compares before, during, and after the Covid-19 pandemic.
- Focus on understanding how to account for BiB4All participants who move outside of Bradford.
- Discuss with health visitors why these data may be missing, to consider how these data could be improved.

These explorations would be meaningful given that ASQ was considered a research priority by local stakeholders, which could generate impetus to improve these data, compared to researchers' wanting to improve these data for their own interest. Collaborations between BaBi teams and local services will be essential for driving forward these improvements.

Based on my experiences of using and analysing linked routine data (Chapter 5), and from speaking to local services throughout my research (including in formal qualitative interviews detailed in Chapter 6), improving the connectivity and quality of NHS data systems would offer considerable benefits for research, evidence, and clinical practice.

In the literature, GPs have described the challenges of communicating patients' needs with health visitors and midwives due to the way routine data are recorded and shared. They described having to use the free text fields to ensure the right information is shared (Pybus *et al.,* 2023, Unpublished). The literature on routine health data has also highlighted the need to collect comparable data across services. Specific examples are described for ethnicity data, where a person may have inconsistent ethnicities recorded across health and care datasets (Scobie *et al*., (2021). Local decision-makers who participated in the qualitative research detailed in Chapter 6, described how these data inconsistencies limit the value of these data for addressing local issues. This could suggest that investment is needed in creating effective data systems, to increase the capability of these data for evidence-based decision-making. This is important as the success of learning health systems is underpinned by high quality data (Department of Health and Social Care, 2022). This type of investment could benefit research conducted with a wide range of NHS data, including early years research.

Issues around routine data collection are also apparent in more established data linkage systems. Brian Dixon, Director of Public Health Informatics at the Regenstrief Institute, which links administrative health data in Indiana, United States, commented that data on social determinants of health are not easily accessible for research. For example, information about if a person has taken time from their low-paying job and taken two bus rides to attend the doctor's appointment, is not typically collected in medical records, or

this is recorded in the free-text field of the medical record. Dixon acknowledges that integrating these data into the medical record, in a way that can be accessed for research, is likely to be a huge task (Regenstrief Institute, 2023). This demonstrates that issues around the quality of routine health data for research are apparent globally, and in more established data linkage systems, as it is challenging to summarise people's experiences and characteristics in a series of codes. The global nature of these issues may facilitate global solutions to improving the use of linked routine data for research, if findings on best practice can be shared across different research groups.

Understanding which clinical codes are most relevant for data analysis was another key challenge. I worked with health and care professionals to understand how clinical codes are used in practice, although, this was not always possible. As such, I was unable to consult health visitors regarding the data collected by their service and I relied on the NHS data dictionary to decide which clinical codes to request. I then worked closely with the BiB data team to determine which data were appropriate for analysis. Based on this, I would recommend a flexible process for researchers to apply for BiB4All data, where researchers can work with the data team to understand what data can be accessed. A single transaction, where a data request is submitted, and all data are provided is unrealistic due to the complex and messy nature of routine data. Hence, adequate funding for the BaBi data teams is necessary to allow them to work with researchers to fulfil data linkage requests in a timely and meaningful manner. Similarly, adequate funding for data teams is likely to apply to other routine data sources which can be used for research purposes. For example, the mapping review identified a study that faced significant delays in accessing these data (Macfarlane *et al.,* 2019).

Moreover, good documentation about how clinical codes are used in practice and the range of values that are recorded could be beneficial, although, developing this documentation is likely to be a huge task and could be the subject of future research. This is because data systems and coding practices vary between NHS Trusts. Data systems also change within each NHS trust over time, which can also impact on how data are recorded.

Finally, researchers could be transparent about their experiences of using linked routine data to provide learning for other researchers planning to use these data. Learning and experiences could be shared by establishing a network of researchers who have or are working with BaBi data, similar to the eCRUSAD (Early Career Researchers Using Scottish Administrative Data) that has been established at The University of Edinburgh (eCRUSADer, 2022). Facilitating an established network of researchers for BaBi data could allow for better use of these data in the future research.

7.1.3 Understanding local decision-makers' perspectives of linked data research

Chapter 6 explored decision-makers perspectives towards linked routine data and using research produced with these data to inform their decision-making. Participants of this research perceived linked routine data as a valuable tool for local early years decision-making. This aligns with the recent large investments made by policymakers in projects that support the use of linked routine data to support children and families (Ministry for Housing, Communities and Local Government, 2021).

Prior to this research, there was a limited understanding of local decision-makers perspectives of linked routine data and why they may not be utilising these data to support early year's decision-making. Despite articulating value in linked routine data, there are possible reasons why limited evidence of these data being used to support early years decision-making was identified in Chapter 2. Thus, the barriers and facilitators to using these data were also explored in Chapter 6 and are the key findings are discussed in the next section.

Participants of the qualitative research also expressed how electronic health record systems influence the quality of routine data, which could limit the value of the data for research and decision-making. This has been discussed above as I also identified these challenges when using data from the BiB4All cohort. The perception that linked routine data are of poor quality for research could explain why limited evidence of linked data informing early years policy and practice was identified in Chapter 2.

The discussions from the qualitative interviews with decision-makers (Chapter 6) revealed that decision-makers would use linked data in conjunction with other evidence in their decision-making. This is consistent with Weiss' (1977, 1979, 1982) theory of enlightenment, which suggests that a body of research influences policy rather than individual research studies. This reaffirms the idea that linked data research may contribute to local decision-making around child and maternal health, alongside other forms of evidence, and that measuring the impact of a single linked data research study on policy and practice can be challenging. This implies that efforts to promote the use of linked data research findings may be worthwhile but measuring the impact of these efforts on research utilisation is likely to be a challenge.

7.1.4 Supporting the use of linked data research by local decision-makers

The barriers and facilitators associated with the use of research by policymakers is widely documented (see section 1.5.1). However, the use of linked data research by policymakers, and the factors influencing this, was not well understood prior to this research (see Chapter 2). I have addressed this knowledge gap by presenting local decision-makers' perspectives towards being engaged and supported to make use of linked data research from the BaBi studies (Chapter 6).

To support the use of linked data research by local decision-makers, my thesis consistently found that developing professional's and communities' understanding of linked routine data, and its potential for research, would be beneficial. For example, educating stakeholders and members of the public on the BaBi data would better enable them to identify priorities for linked data research (see section 7.1.1). Participants of the qualitative research in Chapter 6 recommended educating those that input routine data, during their professional training, about the importance of accurate routine data collection and the implications for research. This could help overcome some of the data quality issues described by participants in the qualitative research, and those I identified in conducing the research in Chapter 5. This could then improve the value of these data for research and decision-making.

However, there may be barriers to improving the way that routine data are collected. For example, professionals may not have the time or tools to correctly enter the information in a way that is relevant for research, meaning data quality and collection issues may persist. Furthermore, health and care professionals are likely to prioritise delivering their services and not gathering data. Nonetheless, good quality data remains essential if it is to be used for research that is relevant to local policy and practice. Bringing together multidisciplinary collaborations of health and care professionals and researchers locally, may enable improvements to routine data systems that allow for better data for both health care delivery and research.

In addition to educating those that input the data, it is important that the data systems that are used to capture routine data are flexible, allowing information that is important locally to be collected. If the data systems do not allow certain information to be recorded or they record data in a way that is not accessible to researchers, then this limits the effectiveness of educating professionals to improve these data for research. This is because professionals are limited to recording information that can be captured by the data system and recording this information in a certain way. For example, clinical

professionals record the ethnicity of a patient using a list of pre-defined categories where they chose the clinical code most appropriate. This links to the discussion in section 7.1.2 regarding the need for an NHS data system that meets the needs of decision-makers (including clinicians, commissioners, and researchers). Further research is needed to identify the barriers to improving the collection of routine data.

The BaBi teams have already established a relationship with the local maternity data system providers and local midwifery teams as part of embedding the BaBi consent process into the routine practice. They have also developed relationships with other local services through their steering groups, which creates the opportunity for BaBi teams to work with these local services and data providers to overcome the issues discussed. It may also be possible to change the data collection systems so that BaBi teams are able to shape how clinicians record data, e.g., changing the system so that the patient record cannot be closed unless particular data fields are filled in. Moreover, there is potential for BaBi teams to work in partnership with those who input routine data to find out how changing the way information is collected would impact their clinical role and what would work best for them. It is important that any changes do not negatively impact on patient care.

The Better Start Programmes have previously demonstrated that it is possible to work with local services to improve the collection of routine data. For example, the Better Start Bradford Innovation Hub began to plug data gaps by working with health visitors to develop a tool for measuring maternal-infant attachment. Small Steps Big Changes in Nottingham worked with early years colleagues and local primary and nursery schools to bridge the gap in missing standardised testing data (The National Lottery Community Fund, 2022). Thus, there is the potential to improve routine data collection in areas setting up BaBi studies.

Developing policymakers' skills in research is identified in the wider *'research use'* literature as supporting the use of research in decision-making (Oliver *et al.,* 2014). Therefore, this research adds to this body of knowledge by identifying the skills needed by decision-makers in areas setting up BaBi studies, to use research that utilises linked routine data.

At an event held by the Nuffield Trust event in November 2019, which is described in Chapter 2, lack of analytical capacity in local authorities was discussed as a barrier to using linked routine data. They also discussed the importance of building an understanding of linked routine data amongst commissioners and incorporating linked data research into their decision-making roles was key. To build this understanding,

participants of the qualitative research in Chapter 6 suggested sharing examples of how linked data have been used by others.

To effectively engage local decision-makers in linked data research, this thesis identified that engagement activities needed to be convenient and involve a short time commitment. This reflects the busy nature of local health and care services. Many of the attendees at the prioritisation workshops in Chapter 4 were able to contribute as the meeting was short (less than two hours) and it was held online. This was reflected in the workshop feedback. When attempting to engage local services in the research conducted in Chapter 5, I found it challenging to find a time when health professionals could meet to discuss the routine data, even when this was a short time commitment. This could be the result of increased pressure on the NHS as a result of the Covid-19 pandemic. In the qualitative research in Chapter 6, participants described how being able to discuss research findings with researchers allowed them to quickly engage with the research findings and make decisions about the evidence. They also wanted the research outputs to be presented as a short summary as this is quick to engage with.

The need for concise, *'short and snappy'* information was emphasised by Hopf *et al*., (2014) in the mapping review in Chapter 2 and is consistent with the wider literature on the use of evidence by decision-makers (Cairney and Oliver, 2018; Dobbins *et al.,* 2007; Lavis *et al.,* (2005); Contandriopoulos *et al.,* 2010; Innvær *et al.* 2002; Mitton *et al.,* 2007; Nutley, *et al.,* 2007; Walter, *et al.,* 2005). Hence, when turning the evidence produced using linked routine data into actionable products for research users (during the translational phase of the research utilisation process, see Figure 5), I recommend representing the research outputs in a way that is quick and convenient for decision-makers to engage with. This could improve the chances of linked data research being used as a local health intelligence tool, alongside other forms of evidence.

There are many initiatives across the world that have been designed to achieve better health through data linkage systems, e.g., Regenstrief Institute, SAIL databank, Data Linkage Western Australia, Manitoba Centre for Health Policy (MCHP) (Regenstrief Institute, 2023; NIHR Applied Research Collaboration South London, 2021; Government of Western Australia Department of Health, 2023; Katz *et al.,* 2021). As the BaBi network is relatively new, it can learn from these more mature data linkage systems elsewhere in the world on how to successfully implement policy. For example, the MCHP have documented the key lessons they have learnt about what makes relationships between policymakers and researchers work (Katz *et al.,* 2021). The MCHP work with the *'Need To Know'* (NTK) team who provide an ongoing platform for two-way communication between researchers and regional health planners in Manitoba. During regular meetings,

health planners share the questions they would like to have answered, and the researchers build capacity among them for understanding how the data can be used to answer these questions. Building capacity among health planners on scientific methods and approaches to using the data supports greater acceptance to the results produced using those data. Researchers also benefit from the insights of planners (Katz *et al.,* 2021). This approach is similar to what was articulated by participants of the qualitative research in Chapter 6, regarding how they wanted to discuss linked data research. Thus, BaBi teams can learn from the work of the MCHP and the NTK team regarding how to create this platform of two-way communication that can support the successful translation of linked data research into policymaking.

Time constraints and lack of capacity were frequently discussed as barriers to engaging with linked data research by participants of the qualitative research in Chapter 6. Many participants described research as not in the nature of their job, when explaining why they do not have time to engage with research. This perception that research is not part of their role creates challenges for engaging these decision-makers in linked data research. This is because these decision-makers may prioritise tasks they perceive to be part of their role over engaging in research. This was apparent when conducting the research in Chapter 5, as I was unable to consult health visitors as part of the data analysis process. They were potentially prioritising their clinical role over engaging with the research. This demonstrates how a decision-maker's perception of who is responsible for research may affect engagement with research. This subsequently impacts how useful the research is to decision-makers and whether it can be used as a local health intelligence tool for child and maternal health.

Therefore, to support the use of linked data by local decision-makers, a change at the systems level may be required, to ensure research is better embedded within decision-maker roles. This could benefit the use of linked data research for health care decision-making as research becomes a more visible priority for decision-makers. A change to the decision-making culture and the way decision-makers' perceive their involvement with research, may also be needed. The introduction of ICSs within the NHS and the HDRCs in local authorities will likely support this change of narrative around the who is responsible for research. Without changing this perception, addressing the other barriers such as lack of skills in linked data research are unlikely to have a significant impact.

My thesis has consistently shown that collaboration between researchers and local services is important when using linked routine data for research and decision-making. The thread of engaging local stakeholders runs throughout my thesis and is important at all stages of the linked data research cycle. For example, Chapter 4 explains why

stakeholder engagement is important in the prioritisation of research, and specifically linked data research. Chapter 2 shows that engagement of local services is a strategy used by data linkage studies to promote the use of their data. This is also consistent with the wider *'research use'* literature, as it is suggested that improving the communications between the policy and research communities can support the use of research by decision-makers (Boswell and Smith, 2017). The importance of actively engaging, liaising, and consulting with key stakeholders to maintain positive working relationships as part of using linked routine data for research is also a key element of *'The Western Australia (WA) data linkage strategy 2022-2024'* (Government of Western Australia, 2023). Moreover, Chapter 5 demonstrated how collaboration between those who collect routine data and those who use routine data for research is essential. It allows researchers to understand how data were collected, so that they can make informed decisions about its reliability for research. This can lead to greater confidence in the research findings and more useful recommendations for policy and practice. It also allows researchers and local services to work together to improve these data. Hence, collaboration between researchers and local stakeholders is essential if linked data are to be used as a local health intelligence tool for child and maternal health.

7.2 Strengths and limitations of the overall thesis

The BaBi Network was an appropriate case study for addressing the aim of this thesis as it allowed an exploration of a number of studies accessing and linking routine data for research across a range of local areas. This means the findings could be useful in several different settings. In addition, the BaBi LHI model provided a useful guide for understanding the challenges faced when using these data for research and translating this into decision-making. For example, there were challenges associated with identifying and addressing local research priorities for linked data research, as well as engaging and supporting decision-makers to use the research evidence. Despite the BaBi LHI model being unpublished, it is underpinned by well-established principles of data-driven decision-making, learning health systems and theories around research utilisation, that allowed me to generate this understanding. I was, however, unable to observe the latter parts of the BaBi LHI model (implementation of a policy change and evaluation of the policy change) within the constraints of this project.

Having attempted to go through the first four stages of this model, I think it would be useful to elaborate on the *'policy or practice'* and *'decision-making'* stages of the BaBi LHI model. A key finding from this research was that existing routine data needs to be improved if it is to be used to inform decision-making. As a result of this, a decision has

been made to carry out further work to explore how ASQ and PMH data are collected locally and how these data can be captured in a meaningful way. This work is a collaboration between researchers at Born in Bradford and local services. Hence, research using these data may not always lead initially to a conventional change to practice, where decision-makers decide to introduce an intervention or change service delivery that would directly influences patient care. It could lead to a change in the way data are recorded by health care professionals, which is the case in this research, that does not initially affect the patient's experience. In the same way that routine data could be used to evaluate the success of an intervention or change to service delivery, we can use these data to explore whether the change to the way these data are recorded has influenced how useful they are for research and decision-making. Hence, I think it is helpful to include this in the model to reflect how researchers can also be part of the decision-making process and how changes to data collection can then lead to improvements in health care. This is important as my research suggests that significant work is needed to improve routine data before it can be used to inform decision-making, and that changes to these data are crucial in moving towards these data being used a local health intelligence tool.

In the FHI 360 Research Utilisation framework (Kim et al., 2018), key activities are detailed at each of the stages, e.g., engaging policymakers is a key activity during the *'Foundational Phase'* of the research utilisation process. As such, it may be helpful to incorporate the learning from this PhD research into the BaBi LHI model to provide guidance at each of the stages. For example, the recommendations from effective engagement of local decision-makers in linked data research, detailed in section 7.1.4, could be included under the *'decision-making'* of the BaBi LHI model. Details of the prioritisation workshop method, detailed in Chapter 4, could be included under the *'Identification and Prioritisation'* stage. This would allow BaBi sites to benefit from the learning generated in this PhD.

Many of the issues identified in this research, such as data missingness and data quality issues, are likely shared with other localities. To share ongoing learning about identified issues, and reduce the duplication of work across sites, there are a number of *'communities of practice'* that exist across the BaBi Network. For example, the data teams at each of the local BaBi sites meet monthly to discuss issues, share practice and to provide updates on their planned work. The BaBi research midwives meet every two weeks to share good practice and provide peer support. The BaBi Network Coordinating Centre share updates with local sites through regular email communication and meetings, which often generates rich discussions in the community of practice meetings. In addition, there is a Google Drive, which collates useful resources for BaBi sites and

records current work being conducted across the Network in an *'impact log'*. Hence, there are various communication channels that allow for collaboration and sharing of knowledge across the Network that can facilitate the use linked routine data for research and decision-making, whilst minimising the potential for duplication of effort across sites.

Using data from the BaBi study in Bradford (BiB4All) to address a research question that was prioritised locally allowed an illustration of the real benefits, challenges, and solutions to using these data to support decision-making. There was also the potential for this research to benefit the local population and to allow BiB to feedback to participants how their data have been used.

A strength of a case study approach is that it can provide in-depth accounts of a phenomena in a real-world setting, although, this may not be generalisable to other settings. As such, further research is needed to understand if the benefits and challenges identified in this research are applicable to other BaBi sites or local areas utilising linked data for research. Moreover, there are many models of linking data for research and the BaBi Network focuses on data linkage at the local level. Thus, some of the findings may not be generalisable to studies linking data on a population or national level.

Another strength is that the teams setting up data linkage studies as part of the BaBi Network, who are the intended audience of the research outputs, were involved throughout this project. They were involved in the design and conduct of the engagement workshops in Chapter 4, and in the planning and recruitment of the qualitative research in Chapter 6. This increases the likelihood that the research outputs will be useful to these teams. This aligns with the applied and practical nature of this research. Moreover, the engagement method that was developed as part of this PhD has formed part of a toolkit to support new and developing BaBi sites, which also now supports the use of this method upon completion of this PhD.

As discussed throughout this thesis, there are a number of different actors that influence the research utilisation process such as journalists, lobbyists, politicians, and researchers. This thesis has considered how researchers and BaBi teams can actively promote the use of linked data research findings. The role of other actors in the use of linked data research was not explored in this thesis. This may be a fruitful line of future research, as this would allow for greater complexity to be considered in the use of linked data research.

Due to delays in receiving the BiB4All data, I conducted the qualitative research detailed in Chapter 6 before conducting the data analysis presented in Chapter 5. This meant that

I was able to draw on the findings from the qualitative research, when conducing and disseminating the BiB4All data analysis. For instance, decision-makers articulated the importance of good quality and robust data to underpin their decision-making. I took this into account when interpreting the results of the data analysis presented in Chapter 5.

Throughout this research, I engaged many local stakeholders, including in the prioritisation of research topics and when analysing BiB4All data. I was transparent about the scope of the involvement and what was to be expected. However, learning from this experience, in future research projects I would consider setting up a stakeholder group to consult throughout the whole project. In addition, the majority of the stakeholders involved in this research, after the research priorities had been identified, were from local health and care services or the local council. In future research, I would like to involve public contributors when analysing the linked data, to ensure their needs are also reflected in the research outputs. This is because many of the considerations I made were based on clinical advice.

A limitation of this thesis is that only routinely collected data from the health service was explored in the data analysis. Therefore, I was unable to comment on the usefulness of routine data from other public services. This could be an important area for future research as much of the data related to the wider determinants of health are captured outside the health service (Sohal, et al., 2022).

Within the scope of this project, I was unable to explore why there were large amounts of missing data for key variables. Section 7.4 details my recommendations for future research.

7.3 Recommendations for BaBi teams

This thesis aimed to produce a set of practical recommendations for new and existing BaBi sites on how to engage and support the use of linked data in their local areas, at each stage of the BaBi LHI model. For example, to recommend a potential method for identifying local research priorities to be addressed using BaBi data (see Chapter 4). These recommendations are aimed at a non-academic audience as the teams managing and setting up BaBi studies are from a range of backgrounds, including midwives, clinicians, and commissioners. These recommendations could also be useful to those working with linked data more generally as the research findings from this thesis align with the wider literature on *'research use'* and linked routine data.

This section presents the key recommendations from this research, in addition to those already discussed throughout this thesis.

In Chapter 1, I presented an impact case study of the GUS research cohort. The activities identified as important for encouraging the use of research by decision-makers in this thesis are consistent with those in the GUS impact assessment. For example, GUS promoted their research findings through a variety of mediums, and this was identified as important by the decisions-makers who took part in the qualitative research in Chapter 6. Hence, to facilitate research impact, I would recommend that the BaBi Network fund a team member who is responsible for the dissemination of their research, such as a dissemination officer. This dissemination officer could track engagement with decision-makers and directly ask policymakers about whether they have made use of the research.

Chapter 2 of this thesis identified a gap in the knowledge for how linked data research has influenced early years decision-making. Therefore, I would also recommend producing a report, similar to the format of the GUS impact assessment, detailing how the research from BaBi has influenced local decision-making and how this was achieved. This would improve the transparency around how linked data research is making a difference to local families, which can support future investments in linked data research. I acknowledge that these recommendations will need substantial funding to implement, and this is not an easy task given the challenges of measuring research impact. It would also involve following up with all the researchers who use BaBi data to find out how their research has informed policy and practice. This is because external researchers can apply to use these data. This becomes even more complex when you consider the use of the BaBi meta-cohort. I would recommend that the BaBi teams ask researchers, as part of the data sharing agreement, to document the impacts of their research and share these with the BaBi team at agreed timepoints.

The process of supporting the use of linked data research in decision-making and tracing this impact could be facilitated by developing partnerships between researchers using linked routine data and decision-makers. This research suggested that creating a space where researchers could present their work and decision-makers could ask questions and discuss potential actions would be beneficial. Researchers could observe how their research is being used to inform decisions and decision-makers can conveniently access the information they need. However, developing these partnerships could be challenging as BaBi data can be accessed by many researchers from different research institutions and for a variety of purposes. Many of these researchers will access BaBi data for a one-off research project. This means that these researchers are unlikely to invest in

developing these relationships. Implementing my second recommendation to track the outputs of these researchers would mitigate some of this impact, as it would allow the BaBi teams to share these findings through their partnerships with decision-makers.

Therefore, I would recommend that the BaBi teams continue to invest in developing partnerships with local decision-makers, to create a platform for researchers who have used their data to easily present their research findings to the relevant decision-makers. The BaBi teams would then be supporting decision-makers to access the local health intelligence they need. This supports researchers to be influential actors within the policy network, and take advantage of the opportunities to influence policy, as discussed in section 1.5.2. Lessons from the MCHP, discussed in section 7.1.4, can be utilised to develop this platform (Katz et al., 2021).

To support researchers to use BaBi data, I would recommend that the BaBi Network provides a list of key information contained in each dataset, standards for data curation, and how these data can be accessed. This could reduce some of the barriers for researchers to use these data for policy relevant research, such as those described in this thesis. This will likely require significant resource to produce and maintain due to the vastness and complexity of the data.

In addition, there is a move towards hosting linked routine data in a TRE, as this can support the development of researchers' skills by allowing researchers to share knowledge of working with these data (Goldacre and Morely, 2022). BaBi could consider hosting their data in a TRE to benefit future use of these data.

These recommendations have been shared with the BaBi teams and made available on the BaBi shared drive for new and existing BaBi sites.

7.4 Recommendations for future research

In addition to the specific recommendations for future research provided within each study, my thesis generated several general recommendations of future work.

This thesis highlighted that missing data was a key challenge for using linked routine data to address local research priorities. ASQ-3 scores were not available for many children in the BiB4All cohort, which is a key outcome variable. It is unclear why these data are missing, and further research is planned to explore if this was because data were unable to be linked for those children; if they weren't recorded due to the Covid-19

pandemic; or if there was another reason. This is important if these data are to be used to conduct policy relevant research.

This research also highlighted the challenges of using routine data to identify the prevalence of PMH. Additional research is planned to explore how these data can be captured in a more meaningful way for research and decision-making. Moreover, further research can also explore whether the issues identified for the BiB4All cohort are the same for other local BaBi cohorts.

This thesis has frequently discussed how educating health and care professionals and decision-makers on the potential value of routine data for research could be beneficial. This could improve the quality of routine data collection for research and the value of the research for decision-making. However, further research is needed to understand how best to achieve this in a way that would allow for better use of these data and that would not detract from their professional role.

The findings from Chapter 2 suggested that further research was needed to understand if linked data research was being used by decision-makers but not documented, or if there were barriers to using these data. This thesis partly addresses this by exploring, at the local level, whether these data had been used previously by decision-makers and the barriers associated with this. Further research is needed to understand this on a broader scale.

In this thesis, I focus on the first four stages of the BaBi LHI model which correspond to the foundational, research and translational phases of FHI 360 Research Utilisation Framework (Kim *et al.,* 2018). Future research could explore the latter stages of the BaBi LHI model or the institutionalisation phase of research utilisation. These stages focus on the change that was made to policy and practice, as a result of the research, and evaluating this change, using the linked routine data, to see if it had an impact on health or practice. This could include exploring how changes made to the collection of routine ASQ and PMH data, as a result of this research, has influenced the utility of these data for research and decision-making. It will be important to understand the challenges and opportunities at each of these stages if BaBi data are to be used as a local health intelligence tool for child and maternal health. Existing research on evaluating policy and practice changes can inform how linked routine data can be used to explore this.

Finally, Chapter 6 of this thesis highlighted that decision-makers' have varying perspectives towards who is responsible for research. Further research could explore whether decision-makers' perceptions towards their role in research influences the time

they are willing to devote to linked data research and its application. This is recommended, given the increasing focus on the role of the health and care workforce in research, particularly within the new ICS structure and the importance of stakeholder engagement in the research utilisation process (NHS, 2023a).

7.5 Recommendations for policy and practice

The findings from this thesis suggest that to realise the benefits of linked routine data from the BaBi studies, it is important that health and care professionals, commissioners, and service providers understand their role in research. As such, I would recommend that clinical practice supports their workforce to accurately collect routine data by developing an understanding around the importance of routine data for research. I would also recommend developing a working culture that is supportive of involvement in research. This will likely be facilitated with the introduction of the new ICS model through their work to develop system-level research strategies, as recommended by NHS England (NHS, 2023a).

It is also important that commissioners fully understand the significance of their role in research and a systems-level approach may be needed to help shape the narrative around who is responsible for research. In Chapter 6, decision-makers suggested that developing ways of embedding the use of linked routine data to inform decision-making in the policy process would support the use of these data and this is a recommendation of this thesis.

7.6 Chapter summary

In this chapter, I have discussed how my thesis has contributed to and added new understanding of how linked routine data, from the BaBi studies, can be used as a local health intelligence tool for child and maternal health. The key finding is that local decision-makers perceive value in linked routine data for early years' decision-making, however, if the potential of these data are to be realised, we need to focus on improving the collection and curation of these data for research. Missing data, and data not capturing the relevant outcomes and exposures, were the key concerns identified in this thesis when attempting to answer one question that was prioritised locally. Therefore, a key recommendation from this research is for BaBi teams to work with local services and data providers to understand more about the way data are captured and opportunities to improve this. This can help ensure key information about families is accurately captured, allowing it to be used to inform decisions that improve their health. This can be supported

through developing partnerships between BaBi teams and local services. How BaBi teams can successfully influence the collection of routine data should be the focus of future research.

This thesis also identified other barriers for decision-makers to engage with and use linked data research. An important barrier related to decision-makers' perspectives towards their responsibility for research. A key priority for research is to explore this further and to understand whether changing the narrative around the responsibility for research, at a systems-level, could lead to greater engagement in linked data research by stakeholders, and how best to do this.

Finally, this thesis highlighted the importance of working with health and care professionals when analysing routinely collected health data, to ensure researchers understand what information these data represent.

The use of existing data or routinely collected data to drive decision-making is high on the policy agenda. As there are several studies being set up with the purposes of linking routine data for research, and there has been significant investment in this area, it is important that we address the issues identified in this thesis, to ensure that we are maximising the benefits of these endeavours.

Appendices

A: Further information for Chapter 2

A1. Search Strategy Pilot - Excluded data sources

Table 54 details the additional sources searched during the pilot that were excluded from the final search strategy and the reasons they were excluded.

Table 54 Excluded data sources from mapping review

Data source	Type of source	Reason for exclusion
Action for children https://www.actionforchildren.org.uk	Children's charity website	No results
Barnardos https://www.barnardos.org.uk	Children's charity website	No relevant results
The Health Foundation https://health.org.uk	Health charity website	No results for data linkage in the topic of children and young people
Early Intervention Foundation https://www.eif.org.uk	Early years charity website	No relevant results
Institute of health equity https://www.instituteofhealthequity.org	Website relating to health equity	No relevant results when searching data linkage search terms
Joseph Rowntree Foundation www.jrf.org.uk	Independent social change organisation working to solve poverty, website	No relevant results to data linkage for early life health.
Nesta https://www.nesta.org.uk	Agency for social good website	One page recommended the use of data linkage to improve data for children's services but was not relevant within the scope of this review. No other relevant results when searching for data linkage
Perinatal Institute https://www.perinatal.org.uk	Not for profit organisation website	no search function or relevant pages
BASE https://www.base-search.net	Database	Lots of irrelevant results

Data source	Type of source	Reason for exclusion
Child link http://www.childlink.co.uk	Database	I do not have access to this site
OpenGrey http://www.opengrey.eu	Database	No relevant articles returned when searching key search terms
Policy commons https://policycommons.net	Database	Subscription required for multiple searches and few relevant articles. It has a limited search function making it difficult to focus the results.
Open DOAR https://v2.sherpa.ac.uk/opendoar/	Database	No relevant results
The Hertfordshire Cohort Study https://www.mrc.soton.ac.uk/herts/	Data linkage study website	No information on findings related to data linkage or child health.
Liverpool Families Programme https://liverpool.gov.uk/children-and-families/liverpool-families-programme/	Data linkage study website	No search function to find key publications.
Maternal and Child Health Network (MatCHNet) http://matchnet.sphsu.gla.ac.uk	Data linkage study website	Website down at the time of the pilot search
National Neonatal Research Database (NNRD) https://www.imperial.ac.uk/neonatal-data-analysis-unit/neonatal-data-analysis-unit/utilising-the-national-neonatal-research-database/	Data linkage study website	Links to outputs not accessible
Next steps, previously known as the Longitudinal Study of Young People in England https://nextstepsstudy.org.uk	Data linkage study website	Children too old
eLIXIR http://www.guysandstthomasbrc.nihr.ac.uk/microsites/elixir/about-the-programme/	Data linkage study website	No research outputs
Millennium Cohort Study (MCS) https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/	Data linkage study website	There was no way of searching for impacts of MCS, but you can look at impacts for the centre for longitudinal studies. There was, however, no results when I searched the search terms. I also navigated to 'publications and resources' and clicked on Briefings and impact. I then clicked 'browse and filter', and selected all document types; early years, childhood and pregnancy, birth and infancy in life stage filters. It is possible to filter which study you are interested in out of MCS, National Child Development Study (NCDS) and 1970 British Cohort study (BCS70).

Data source	Type of source	Reason for exclusion
		I included all within the search filter except next steps as the children are too old. I did filter on each of the different studies to ensure the children were below the age of five such as NCDS birth sweep, BCS70 birth sweep and age five sweep. There were no relevant results.
Northern Ireland Longitudinal Study (NILS) https://www.nils-rsu.co.uk	Data linkage study website	Outputs found in CALLS-HUB
Office for National Statistics Longitudinal Study (ONS LS) https://www.ons.gov.uk/aboutus/whatwedo /paidservices/longitudinalstudyls	Data linkage study website	Outputs found in CALLS-HUB
Scottish Longitudinal Study (SLS) https://sls.lscs.ac.uk	Data linkage study website	Outputs found in CALLS-HUB
Qresearch https://www.qresearch.org	Data linkage study website	Browsed the first 100 papers in the publications section and none related to early life health
Scottish Informatics and Linkage Collaboration (SILC) https://www.datalinkagescotland.co.uk-	Data linkage study website	Findings of SILC were difficult to search, it is likely that published studies will mention where data was provided by SILC.
Scottish Health Informatics Programme (SHIP) has moved to FARR institute https://www.farrinstitute.org	Data linkage study website	No results in publications for data linkage
Southampton Women's Survey (SWS)	Data linkage study website	No relevant results
Think Family Database (TFD) by Bristol insights group ▪ https://www.bristol.gov.uk/policies-plans-strategies/the-troubled-families-scheme ▪ https://insightbristol.wixsite.com/home/think-family-database	Data linkage study website	Is not easily searched for impact or research findings.
Health Data Research UK (HDR UK) https://www.hdruk.ac.uk	Data linkage study website	Research not focused on child and early life health.
National Institute for Health and Care Excellence (NICE) https://www.nice.org.uk	Organisation that provides guidance to improve health and social care website	No relevant results to data linkage on the website.
Ministry of Justice in gov.uk	Policy website known to link to health data	Checked the first 30 results when searched for data linkage and none related to health

Data source	Type of source	Reason for exclusion
eDRIS https://www.isdscotland.org/Products-and-Services/EDRIS/	Data service	Navigated to 'publications' page of the website and filtered by health topic to narrow down to child health. The results did not appear to link to data linkage.
Research Excellence Framework https://www.ref.ac.uk	System for recording the impact of research	I searched terms "linkage", "linked data", "Routine data" and "administrative data", where in each case I filtered on UK results. "linkage, "administrative data" and "linked data" searches had no results that were relevant to child health. "Routine data" search identified one potentially relevant article but the use of linked data was unclear.

A2. Final Search Strategy

Table 55 presents the strategy used to search Medline via Ovid. All other database searches can be found on the online open access repository, Figshare (https://doi.org/10.6084/m9.figshare.24005949.v1). This link also provides details of all grey literature searches.

Table 55 Medline via Ovid 1996- week 5 2021

Data searched: 12/10/2021

Term	Results
1. “Early years”.tw 2. Infant.tw 3. Infancy.tw 4. Baby.tw 5. Babies.tw 6. Toddler*.tw 7. Preschool.tw 8. “First years”.tw 9. “Early childhood”.tw 10. Child*.tw 11. Antenatal.tw 12. Postnatal.tw 13. Newborn.tw 14. Mother*.tw 15. Parent*.tw 16. Family.tw 17. Families.tw 18. Maternity.tw 19. Maternal.tw 20. Father.tw 21. Paternal.tw 22. Pregnancy.tw 23. Pregnant.tw 24. Perinatal tw 25. Paediatric.tw 48. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25	2935 101468 32613 25222 23606 9353 17319 2961 20291 923132 27386 76422 62205 154315 293125 613777 186705 14992 189091 16317 17611 243849 126067 50767 47851 2197982
26. “Data link*”.ti 27.“Link* adj3 data*”.ti 28. “Integrated data*”.ti 29. “Connected data”.ti 30. “Record link*”.ti 31. “information link*”.ti 32. “linked electronic health record*”.ti 33. “electronic birth cohort”.ti 34. e-cohort.ti 49. 26 OR 27 OR 28 OR 29 OR 30 OR 31 OR 32 OR 33 OR 34	503 1229 288 0 698 13 21 4 27 2257
35. UK.tw 36. United Kingdom.tw 37. Great Britain.tw 38. England.tw	95020 30427 4173 37284

Term	Results
39. Scotland.tw 40. Wales.tw 41. Northern Ireland.tw 42. Welsh.tw 43. Scottish.tw 44. English.tw 45. Britain.tw 46. British.tw 47. “New South Wales”.tw 50. 35 OR 36 OR 37 OR 38 OR 39 OR 40 OR 41 OR 42 OR 43 NOT 44 OR 45 OR 46 OR 47	11898 17320 3622 1369 6736 84019 7904 31778 7108 271490
48 AND 49 AND 50	112

B: Further information for Chapter 4

B1. Guidance document for prioritisation workshop facilitators

Agenda for the workshop

10:00	**Join the Zoom call**
10:05	**Welcome and Introductions** (Chair- 10 mins) *Welcome attendees to the child and maternal health prioritisation workshop.* *The purpose of the session is to is to bring people together to discuss important areas of child and maternal health that can be explored with BaBi data.* *Remind attendees that the session is going to be recorded.* (Technical support facilitator starts the recording)
10:10	**Ground Rules** (Chair- 5 mins) See ground rules section below
10:15	**Ice Breaker** (Chair- 5 mins) *Attendees will be asked to turn their cameras off and turn them back on to wave if the category applies to them. E.g., wave if you are a parent… a researcher…etc.*
10:20	**Background presentation** *A member of the team will briefly talk though what BaBi is and how we got to this workshop.*
10:30	**Opportunity to ask questions** (Chair- 5 mins) *See if attendees have any questions. A member of the facilitation team to assist in checking the chat for questions.*
10:35	**Explain first group task** (Chair-5 mins) *Explain that attendees will now be placed into small groups for 20 minutes and asked to discuss areas of child and maternal health that they want to know more about and that they think could be answered with the data we have described. Explain that one member from each group will feedback the main points from the discussion to the main group.* (Technical support facilitator initiates the breakout rooms for session one). (Chair - ask attendees to join the breakout room)
10:40	**Breakout session one** (All facilitators) Please see below Breakout Session one
11:00	**Whole group feedback** (Chair- 15mins) *Ask a member from each group to feedback their key ideas. Ask the public contributor's group to feedback first and ask who the nominated group member was, followed by Midwives, Health visitors, Clinicians, and Commissioners.* *Once each group has fed back explain that after the break the groups will be mixed up and they will be placed into new groups. They will be asked to prioritise the research areas that were generated in the first task, based on urgency and importance. One member from each group will then feedback the group's favourite idea to the main group.*

11:15	**Short comfort break** (Chair - 10 mins)
11:25	**Return from the break** (Chair) *Ask if attendees can turn their cameras on so facilitators can see if everyone is back from the break.* (Technical support facilitator initiates breakout rooms for session two once people have returned from the break).
11:28	**Breakout session two** (All facilitators) Please see below Breakout Session two.
11:45	**Whole group feedback** *A member from each group to feedback their most important research idea (30 seconds per group). A brief discussion reflecting on how members found the session.*
11:55	**Concluding remarks** (Chair- 5mins) *Thank everyone for attending the session and ask if anyone has any final remarks they would like to make.* *Explain how the outputs from the session will be used to shape our research in BaBi going forwards. The outputs will also be shared with the Early Life and Prevention theme of the Yorkshire and Humber Applied Research Collaboration to feed into the research priorities being identified across the region.* *Inform attendees that a feedback form will be sent out after the session, and we would be really grateful if they could fill it out.*
12:00	**End the session**

Ground rules (All facilitators in their breakout session)

a) We would like you to share your views and questions in a respectful and tolerant way and encourage a discussion.
b) We appreciate that every participant brings different expertise, experience, and perspective, and all inputs are valuable.
c) All participants will be given equal opportunity to express their views.
d) Please can we maintain confidentiality for those in the group by not sharing any of the details discussed in this workshop, outside of this group.
e) We will aim to start and end the session on time.
f) To help ensure the workshop runs smoothly,
    (i) Please can you mute your microphone unless you are speaking.
    (ii) We will listen when others speak and avoid interrupting them so that only one conversation happens at a time.
    (iii) In the breakout rooms, we would encourage you turn your cameras on, if you are willing and able, as will help with the discussion.
    (iv) Given the number of participants, please use the chat function or raise your hand to ask a question in the main group sessions.

g) If you have any technical issues and get disconnected from the session, just use the same link we provided in the joining email, to re-join. If you have any problems re-joining email *name*, their details are provided in the chat.

h) You can stop being involved at any point, just message *name* or another facilitator to tell them that you are leaving the workshop so that we know you are not re-joining the call and then feel free to exit the zoom call.

i) Please feel free to ask any questions before we begin or at any point during the presentation through the chat function. You can raise your hand if you would like to speak.

Breakout room one (20 mins)

a) Introduce yourself and notetaker.

b) Invite attendees to briefly share name, where they are from, and their role or reason for joining the workshop.

c) Outline to attendees how you would like them to contribute to the discussion.

d) Share your screen to show the Google Jamboard for session one.

e) The first slide will be used to notes down attendee ideas on sticky notes. Please click on the icon highlighted in the diagram below to add a sticky note.

f) The second slide, which can be accessed by clicking the arrow at the top as above, shows a diagram with the datasets with consent to be linked as part of the study. This can be used to remind participants to think of topics that can take advantage more than one data source.

g) Please use the colour sticky note that is the same colour as the title on the board to note down contributors' ideas.

h) On each sticky note, please make it clear what the direction of the relationship, such as which variable is the outcome variable, and which is the exposure variable. This will make it easier for other facilitators to interpret in the following session.
i) Reiterate there are no right or wrong answers and that every contribution is valuable.
j) Establish a group member to feedback one idea from the group to the main group.
k) Might be useful to define important as something that will likely have a large impact on the community and significant to improving child and maternal health outcomes and urgent as something that is time dependant and requires immediate action.
l) There will be a timer in the corner of the breakout room for you to check how much time is left for discussion.

The aim of the session is to generate between six and eight sticky notes with ideas for research relating to maternal health, child health, development and wellbeing that could be explored using BaBi data.

Could start off the session by asking the group to think about one thing that they think we should be focusing on with this data, or one area of child and maternal health they think is important to be researching now. Proceed by expanding this idea to achieve the relationship between two variables.

If after ten minutes you already have six ideas, encourage more detailed discussion around each and use this time to focus on developing them as research questions and prioritising their key ideas. If people are struggling to come up with an idea, focus on one area that has been mentioned, or to think about one area of early life health that is important for future development. Encourage the group to think about what would be interesting to investigate surrounding this topic, with reference to the data we have access to (e.g., maternity, health visiting, primary care records, school records, etc).

As prompts, the data can be used to:

1) Describe a problem and the extent to which is exists. This could include health problems themselves or access to health services.
2) Design and/or evaluate an intervention to tackle an already identified problem.
3) Evaluate the impact of a change to the provision of a service.

If there are no suggestions at all, please refer to the child and maternal health Public Health Outcomes Framework for some potential discussion areas.

Breakout Room two (15 – 20 mins)

a) Introduce yourself and notetaker
b) Invite attendees to briefly share their name, role or reason for joining the workshop.
c) Outline to attendees how you would like them to contribute to the discussion.
d) Open the Jamboard for breakout session two and share screen with attendees.
e) You will now have a selection of sticky notes that were generated in the first session which you can use your mouse to drag around the board.
f) Establish a group member to feedback one idea into the main session in 30 seconds.
g) You are free to ask if there are any ideas, they have thought about over the break, that they would want to add.

The aim of this session is to prioritise the ideas using the urgency and importance matrix. The definitions of what is meant by important and urgent will be on the slide (or refer to the definitions provided in Breakout room one).

h) Emphasise that just because they don't rank something high in both, that doesn't mean that it is not important, you want them to base it on the definitions we have provided.
i) Discuss reasons why they are placing each idea in that section of the matrix and try to reach an agreement on the ideas.
j) Could start the discussion by asking whether each idea makes sense and the proceed with the ideas they think are most important and urgent.

B2. Email invitation for prioritisation workshop

Hi,

The Born in Bradford research team would like to invite you to a workshop to help us decide how best to make use of a linked dataset that forms part of an exciting new electronic birth cohort study in Bradford. The cohort links broad ranging routinely collected data across health, social care, and education sectors. The workshop aims to generate and prioritise research questions that are relevant to local services and families.

The workshop will take place online using Zoom and we are inviting health professionals, service providers, commissioners, and parents to attend. **The workshop will take place on (date/time).**

You don't need to have any special knowledge about Born in Bradford or about research to take part. All the information you need will be explained in the workshop or shared in advance.

If you are interested in taking part in the workshop, please register here. If you have any questions about the workshop, please contact: *contact details*

We hope to see you at the workshop.

B3. Information sheet for prioritisation workshop

Born in Bradford engagement workshop 18th March 2021

Background

BiB4All is an exciting new electronic birth cohort study that asks women during their pregnancy for their permission to access and use information that is routinely collected about themselves and their child for research purposes. BiB4All aims to harness the power of routinely collected data from multiple services and organisations to build a clearer picture of children's and families lives over time. Once connected, the data will be used by researchers, in collaboration with local services and commissioners, to help improve health, care and services through research and planning.

We have set up the project in Bradford which has shown that it is feasible and acceptable to link data across maternity, health visiting, GP, laboratory, social care, and education services. Our ambition is to create a series of electronic cohorts (Born and Bred In (BABI)) across the country that can be used locally as health intelligence tools to shape services and brought together as one to answer research questions of national relevance. We are currently working with teams in Leeds, Doncaster, and Wakefield to set up the project in their area.

We believe this is a ground-breaking idea that can advance science, improve the lives of a new generation, and engage families in the UK in the power of data to shape our future.

What to expect from the workshop

The aim of the workshop is to bring people together to discuss important areas of child health and development that can be explored using the data now and in the future. The outcomes from this workshop will inform the research carried out as part of the BiB4All study and similar studies across the region.

The workshop will be hosted on Zoom and last two hours, with a break after one hour. The workshop will begin with introductions from the team and a short presentation explaining the study in more detail. You will then have the opportunity to ask any questions you may have. Workshop attendees will be from a variety of backgrounds including parents, midwives, health visitors, clinicians, service commissioners, researchers and members of the community who work with or represent parents.

There will be two main sessions in the workshop. First, you will be placed into small groups with people from a similar working background where you will be asked to discuss areas of child health and development you think are important to know more about and could be explored with the linked data from the project.

For the second session, you will be placed into new groups, so that you are speaking to people from different services. You will be asked to think about the suggestions put forward by the groups in the first session and try to prioritise them. Below is an outline agenda for the workshop.

10:00	Join the Zoom call
10:05	Introduction to the session and housekeeping
10:20	Short background presentation
10:30	Opportunity to ask questions
10:35	First group task Discuss areas of child health and development that could be explored with the data.
10:55	Whole group feedback A member from each group to feedback their key ideas
11:15	Short comfort break
11:25	Second group task Prioritise the research ideas from the first exercise.
11:40	Whole group feedback A member from each group to feedback their most important research idea. A brief discussion reflecting on how members found the session.
11:55	Concluding remarks
12:00	End the session

C: Further information for Chapter 5

C1. Email Invitations for additional consultation

C1.1 Email Invitations for public contributors

Hi,

The Born in Bradford research team would like to invite you to attend an online session to help us to decide on a research question to be addressed with data from our exciting new birth cohort study, BiB4All.

In March 2021, the research team organised a workshop which brought together members of the public and local health professionals to discuss areas around child and maternal health they thought were important and urgent for research. In this session, you will be choosing which of these questions we should be addressed as part of a PhD project.

The session will take place online using Zoom and we are inviting parents and health professionals. **The session will take place on 5th March 1-3pm** but should last no longer than one hour forty-five minutes.

You don't need to have any special knowledge about Born in Bradford or about research to take part. All of the information you need will be explained in the session or shared in advance. You can also have a call with someone from the team before the workshop to ask any questions you have.

As a thank you for your time, you will receive £25 for taking part. If you are interested in taking part, please email Hollie Henderson at *email address* who is arranging the session. Hollie will then send you a link to join. If you have any questions, please also email Hollie.

We hope to see you there.

C1.2 Email Invitations for service providers and commissioners

Hi,

The Born in Bradford research team would like to invite you to attend an online session to help us to decide on a research question to be addressed with data from our exciting new birth cohort study, BiB4All. The cohort links broad ranging routinely collected data across health, social care, and education sectors.

In March 2021, the research team organised a workshop which brought together members of the public, local health professionals and commissioners to discuss areas around child and maternal health they thought were important and urgent for research as part of the BiB4All project. The aim of this session is to prioritise one of these research questions, which will then be addressed as part of a PhD project.

The session will take place online using Zoom and we are inviting health professionals, child and maternal health service providers, commissioners, and parents to attend. **The session will take place on 5th March 1-3pm** but should last no longer than one hour forty-five minutes.
You don't need to have any special knowledge about Born in Bradford or about research to take part. All the information you need will be explained in the session or shared in advance.

If you are interested in taking part, please email Hollie Henderson at *email address* who is arranging the session. If you have any questions, please also email Hollie.

We hope to see you there.

C2. Additional consultation background information sheet

Born in Bradford Prioritisation Session 29th March 2022 Background

Born in Bradford for All (BiB4All) is an exciting new electronic birth cohort study that asks women during their pregnancy for their permission to access and use information that is routinely collected about themselves and their child for research purposes. This allows for data such as health records, social care, and education to be linked for those individuals, building a clearer picture of children's and families lives over time. Once connected, the data will be used by researchers, in collaboration with local services and commissioners, to help improve health, care and services through research and planning. For more information about BiB4All please watch the following video: https://www.youtube.com/watch?v=aPkYOlHBV_E.

Our ambition is to create a series of electronic cohorts (Born and Bred In (BABI)) across the country that can be used locally as health intelligence tools to shape services and brought together as one to answer research questions of national relevance. We are currently working with teams in Leeds, Doncaster, and Wakefield to set up the project in their area.

What to expect from the session

In March 2021, we hosted an online two-hour stakeholder engagement workshop over Zoom. The aim of the workshop was to bring people together to discuss important areas of child and maternal health that can be explored using the BiB4All data now and in the future. Workshop attendees were from Bradford, Leeds, Doncaster, Wakefield, and Sheffield and represented a variety of backgrounds including parents, midwives, health visitors, clinicians, service commissioners, researchers and members of the community who work with or represent parents.

During the workshop, attendees generated a list of ideas for research and then prioritised these ideas. Ideas were prioritised based on urgency, defined as something time dependent and requires immediate action, and importance defined as something that will likely have a large impact on the community and significant to improving child and maternal health outcomes. This resulted in a list of 17 topics being identified as both

important and urgent. Additional workshops have been hosted in other local areas including Doncaster, where we have combined the outputs from across these sessions. The research ideas from across the sessions have been narrowed down into those that have the potential to be addressed with linked routine data from BiB4All.

The aim of this session is to prioritise one of these ideas for research now with the BiB4All data. We want you to help us decide which of these ideas is most relevant to our local population in Bradford. The outputs from this session will directly inform the research question addressed as part of a PhD project. The purpose of this is to ensure the research questions addressed as part of BiB4All are relevant to the challenges faced locally in Bradford.

The session will be hosted on Zoom and last approximately 1 hour 45 minutes, with a comfort break after 45 minutes. We are inviting health professionals, child and maternal health service providers, commissioners, and parents from Bradford to attend.

The session will begin with introductions from the team and those attending, followed by a short presentation explaining the study in more detail. You will then have the opportunity to ask any questions you may have. There will be two main parts of the session. The first part will involve a discussion around the prioritised ideas, to ensure we are clear on what each of the ideas represent. The second part will involve a prioritisation task, where you will be asked to decide which idea is the highest priority based on a set of criteria.

C3. Feedback on additional consultation with stakeholders

Four attendees completed the feedback form using Google forms, where the feedback received was positive. Attendees who completed the form felt the session was well organised, they had enough opportunity to contribute to the discussion and enjoyed the session. One attendee expressed that the information sheet they received prior to the session describing what to expect was *"very helpful".* Two attendees felt that more participants might have improved the session. However, the inclusion of more participants would have meant extending the time of the session which would result in lower attendance. This was also not possible due to the extra pressures faced by the NHS at this time. Attendees were asked to provide feedback on how they found the first task of defining the research questions and rate this on a scale of one to five, where *'5'* was challenging and *'1'* was easy. There were two responses that rated this task a *'3'*, one attendee found the task easy, and one attendee rated this task a *'4'*. This suggests that developing contributors' skills in research may benefit future engagement. Finally, one attendee expressed their preference for an in-person meeting, which can be considered for future engagement.

C4. BiB4All data request

Table 56 shows the datasets that were provided by the data team for analysis in Chapter 5 after multiple iterations. It details the source of these data, the codes and terms included and whether these data related to the mother or the child participant. All datasets included Person ID to identify which participant the data related to. The Birth dataset from the maternity data warehouse included both mother and child Person IDs. The data team were able to the 'delivery serial number' in the Birth and Delivery datasets to link add the mother IDs into this table.

Table 56 BiB4All datasets cleaned, linked, and analysed in Chapter 5

Dataset	Data source	Codes/Terms applied for and the associated description	Mother or child data
BiB4All_ASQ_HV	General Practice and Health Visiting	(CTV3 code) CTV3 term: (XacDr) ASQ-3 12 month questionnaire- communication score; (XacDs) ASQ-3 12 month questionnaire - fine motor score; (XacDt) ASQ-3 12 month questionnaire - gross motor score; (XacDu) ASQ-3 12 month questionnaire - personal-social score; (XacDv) ASQ-3 12 month questionnaire - problem solving score; (XacEL) ASQ-3 24 month questionnaire - communication score; (XacEM) ASQ-3 24 month questionnaire - fine motor score; (XacEN) ASQ-3 24 month questionnaire - gross motor score; (XacEP) ASQ-3 24 month questionnaire - personal-social score; (XacEO) ASQ-3 24 month questionnaire - problem solving score; (XacEQ) ASQ-3 27 month questionnaire - communication score; (XacER) ASQ-3 27 month questionnaire - fine motor score; (XacES) ASQ-3 27 month questionnaire - gross motor score; (XacEU) ASQ-3 27 month questionnaire - personal-social score; (XacET) ASQ-3 27 month questionnaire - problem solving score; (XacEV) ASQ-3 30 month questionnaire - communication score; (XacEW) ASQ-3 30 month questionnaire - fine motor score; (XacEX) ASQ-3 30 month questionnaire - gross motor score; (XacEZ) ASQ-3 30 month questionnaire - personal-social score; (XacEY) ASQ-3 30 month questionnaire - problem solving score	Child

Dataset	Data source	Codes/Terms applied for and the associated description	Mother or child data
		Value associated with these codes and terms. **Other fields:** Date code was recorded Age_years Age_months	
BiB4All_GP_PMH **BiB4All_HV_PMH**	General Practice and Health Visiting	The same CTV-3 codes were searched in both the General Practice and Health Visiting datasets. There are more than 500 codes, and these can be made available on request.	
BiB4All_Birth	Maternity data warehouse – Birth table	delivery_method_current_baby baby_breast_milk_status_at_discharge_from_hospital baby_breast_milk_status_at_discharge_from_hospital_desc baby_ffeed_brstmlk baby_ffed_brstmlk_desc birth_order_maternity_services delivery_method_current_baby pregnancy_outcome_fetus baby_year_of_birth baby_month_of_year	Child
BiB4All_Breastfeeding_GP **BiB4All_Breastfeeding_HV** **BiB4All_Breastfeeding_CH**	General Practice, Health Visiting and Healthy Child Programme	The same CTV-3 codes were searched in each data source and provided in the resulting datasets. There are more than 100 clinical codes and these can be made available on request.	Mother and Child

Dataset	Data source	Codes/Terms applied for and the associated description	Mother or child data
BiB4All_delivery	Maternity data warehouse- Delivery table	gestation_length_in_days_dating_ultrasound_scan gestation_length_in_weeks_at_assessment labour_or_delivery_onset_method_code labour_or_delivery_onset_method_code_desc delivery_month delivery_year	Mother
BiB4All_critical_care_neonatal	Maternity warehouse data	critical_care_length_of_stay	Child
BiB4All_child_sex	Maternity warehouse- Live Patient table	sex	Child
BiB4All_GP_ethnicity	General Practice	**(CTV3 code) CTV3 term:** (XaJQv) British or mixed British - ethnic category 2001, (XaJQw) Irish - ethnic category 2001 census, (XaJQx) Other White background - ethnic category 2001, (XaJQy) White and Black Caribbean - ethnic category 2001 census, (XaJQz) White and Black African - ethnic category 2001, (XaJR0) White and Asian - ethnic category 2001 census, (XaJR1) Other Mixed background - ethnic category 2001, (XaJR2) Indian or British Indian - ethnic category 2001, (XaJR3) Pakistani or British Pakistani - ethnic category 2001 census, (XaJR4) Bangladeshi or British Bangladeshi - ethn categ 2001 census, (XaJR5) Other Asian background - ethnic category 2001, (XaJR6) Caribbean - ethnic category 2001 census, (XaJR7) African - ethnic category 2001 census, (XaJR8) Other Black background - ethnic category 2001, (XaJR9) Chinese - ethnic category 2001 census, (XaJSg) Any other group - ethnic category	Mother

Dataset	Data source	Codes/Terms applied for and the associated description	Mother or child data
		2001 census, (XaJRB) Ethnic category not stated - 2001 census, (XaE4B) Ethnic group not given - patient refused. **Other fields:** Date code was recorded Age_years Age_months	
BiB4All_IDs_dates	Maternity data warehouse	ConsentDate Expected_delivery_date_year Expected_delivery_date _month delivery_year_consent_table delivery_month_consent_table	Mother
BiB4All_IMD_variables	Generated by the BiB data team	IMD_2010_decile IMD_2010_score IMD_2019_decile IMD_2019_score	Mother
BiB4All_Patient	Maternity data warehouse – Patient table	Ethnic_origin	Mother
BiB4All_SystmOne_Matched_Indicator		systmone_matched (1/0 values)	Mother and Child

C5. Sensitivity analysis for ASQ 3 cut off values

To understand whether the choice of the cut-off value for a child being considered 'at risk of developmental delay' using the ASQ score affects the results, I conducted sensitivity analyses using cut-off values 10% below and above the cut-off value used in the analyses, where the mean ASQ score was close to the cut-off. The results are presented in Table 57 and Table 58.

Table 57 shows that for cut-off points 10% above those used in clinical practice, there are no significant associations between an indication of PMH and risk of developmental delays for the problem solving gross motor and fine motor domains of the ASQ-3 24 months. The odds of developmental delay for problem solving (OR=1.97, CI=0.61 to 6.37 p=0.26) gross motor (OR=2.66, CI= 0.82 to 8.60, p=0.10) and fine motor (OR=3.07, CI=0.77 to 12.26, p=0.11) domains of the ASQ-3 24 month questionnaire for children of mothers experiencing poor PMH were greater than children of women who did not have an indication of poor PMH, although these odds were smaller than those included in found in Table 41. Poor PMH is also no longer significantly associated with risk of developmental delay in the gross motor domain of the ASQ-3.

Table 58 shows that for cut-off points 10% below those used in clinical practice, an indication of PMH is not associated with risk of developmental delay for the problem solving (OR=3.73, CI=0.99 to 14.12, p=0.05), gross motor (OR=3.39, CI=0.87 to 13.25 p=0.08) and fine motor (OR=5.65, CI=0.75 to 42.65, p=0.09) domains of the ASQ-3 24 months. The odds of risk to developmental delay for the gross motor and fine motor domains of the ASQ-3 24 months for children of women with an indication of poor PMH have increased, although these are not statistically significant. This demonstrates that the results are sensitive to the clinical cut-off points. Small sample sizes and large confidence intervals suggest uncertainty around these estimates.

Table 57 Logistic regression analysis estimating the odds ratio for the risk of developmental delay for children aged 24 months for ASQ cut offs 10% above those used in clinical practice

ASQ	Problem solving (N=84)		Gross motor (N=84)		Fine motor (N=84)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH						
Indication of PMH	1.97 (0.61 to 6.37)	0.26	2.66 (0.82 to 8.60)	0.10	3.07 (0.77 to 12.26)	0.11
IMD decile 2019						
2	0.38 (0.09 to 1.57)	0.18	0.58 (0.15 to 2.26)	0.43	0.08 (0.01 to 0.50)	0.01*
3	1.31 (0.21 to 8.40)	0.77	0.64 (0.10 to 4.07)	0.64	0.16 (0.02 to 1.49)	0.11
4	1.31 (0.17 to 10.33)	0.80	0.14 (0.01 to 1.53)	0.11	4.25 (0.33 to 54.16)	0.27
5	3.08 (0.34 to 28.03)	0.32	2.18 (0.24 to 20.05)	0.49	1.01 (0.08 to 12.58)	0.99
6	3.06 (0.14 to 66.30)	0.14	2.71 (0.13 to 57.11)	0.52	0.31 (0.02 to 5.89)	0.44
7	-	-	-	-	-	-
8	1.00	-	-	-	-	-
9	-	-	-	-	-	-
10	-	-	-	-	-	-
Maternal age						
<24 years	2.66 (0.69 to 10.20)	0.15	3.09 (0.81 to 11.79)	0.10	2.49 (0.58 to 10.63)	0.22
>35 years	1.08 (0.21 to 5.53)	0.93	0.64 (0.13 to 3.16)	0.58	0.83 (0.14 to 4.82)	0.84
Gestational age						

ASQ	Problem solving (N=84)		Gross motor (N=84)		Fine motor (N=84)	
<37 weeks	0.29 (0.03 to 2.43)	0.25	0.39 (0.05 to 3.12)	0.37	1.30 (0.15 to 11.28)	0.81
Mode of delivery						
Elective caesarean section	6.31 (1.41 to 28.27)	0.02*	4.01 (0.95 to 17.03)	0.06	7.99 (1.51 to 42.34)	0.02*
Emergency caesarean section	6.95 (0.63 to 77.05)	0.11	1.81 (0.24 to 13.47)	0.56	1.80 (0.22 to 14.77)	0.59
Other	3.86 (0.88 to 16.98)	0.07	3.51 (0.74 to 16.74)	0.12	3.25 (0.62 to 17.03)	0.16
NICU						
NICU admission	2.85 (0.64 to 12.70)	0.17	2.10 (0.51 to 8.69)	0.31	5.38 (1.08 to 26.87)	0.04*

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level.
OR = Odds Ratio
CI= Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

Table 58 Logistic regression analysis estimating the odds ratio for the risk of developmental delay for children aged 24 months for ASQ cut offs 10% below those used in clinical practice

ASQ	Problem solving (N=81)		Gross motor (N=84)		Fine motor (N=76)	
	OR (CI)	P value	OR (CI)	P value	OR (CI)	P value
PMH						
Indication of PMH	3.73 (0.99 to 14.12)	0.05	3.39 (0.87 to 13.25)	0.08	5.65 (0.75 to 42.65)	0.09
IMD decile 2019						
2	0.55 (0.13 to 2.37)	0.42	0.21 (0.04 to 1.10)	0.07	0.06 (<0.01 to 1.02)	0.05
3	1.96 (0.30 to 12.85)	0.48	0.38 (0.05 to 2.81)	0.34	0.40 (<0.01 to 0.74)	0.03*
4	1.31 (0.16 to 10.84)	0.80	0.27 (0.02 to 3.03)	0.29	1.00	-
5	4.87 (0.48 to 49.24)	0.18	1.07 (0.09 to 13.40)	0.96	2.32 (0.16 to 34.25)	0.54
6	1.00	-	0.58 (0.04 to 9.37)	0.70	1.00	-
7	-	-	-	-	-	-
8	1.00	-	-	-	1.00	-
9	-	-	-	-	-	-
10	-	-	-	-	-	-
Maternal age						
<24 years	1.42 (0.38 to 5.34)	0.61	2.16 (0.56 to 8.35)	0.26	0.27 (0.04 to 1.85)	0.18

ASQ	Problem solving (N=81)		Gross motor (N=84)		Fine motor (N=76)	
>35 years	0.47 (0.08 to 2.69)	0.40	0.31 (0.05 to 1.88)	0.20	0.07 (<0.01 to 1.08)	0.06
Gestational age						
<37 weeks	0.92 (0.11 to 7.81)	0.94	0.81 (0.10 to 6.76)	0.85	7.03 (0.57 to 87.03)	0.13
Mode of delivery						
Elective caesarean section	3.99 (0.92 to 17.32)	0.07	3.42 (0.76 to 15.35)	0.11	56.31 (4.27 to 743.29)	<0.01*
Emergency caesarean section	3.14 (0.42 to 23.68)	0.27	0.91 (0.12 to 6.85)	0.92	1.69 (0.20 to 14.51)	0.64
Other	3.27 (0.77 to 13.86)	0.11	3.51 (0.75 to 16.44)	0.11	4.21 (0.60 to 29.41)	0.15
NICU						
NICU admission	3.42 (0.80 to 14.70)	0.10	4.74 (1.05 to 21.33)	0.04*	14.08 (1.80 to 110.18)	0.01*

Base factors excluded from the regression: (any PMH) no PMH, (IMD decile 2019) 1, (Maternal age) 25-34 years, (GA) ≥ 37 weeks, (Mode of delivery) normal or cephalic vaginal delivery, (NICU) no NICU admission.
Asterisk indicates statistical significance at the 5% level.
OR = Odds Ratio
CI= Confidence Interval
Those with no observations are denoted with '-'
Those with small sample sizes and no variation within have no CI.

D. Further information for Chapter 6

D1. Qualitative interview email invitation

Dear Colleague,

Hollie Henderson, a PhD student at the University of York working with Born in Bradford, would like to invite you to take part in an online interview to share your views on linked data and its potential to inform decision-making related to perinatal and early life health. As someone working in the field of maternity, perinatal and early life health, your views are extremely valuable in learning about how we can support you to make use of the research produced as part of an exciting new programme of electronic birth cohort studies across the region. These electronic birth cohorts link broad ranging routinely collected data across health, social care and education sectors. This research will help develop a set of tools to engage and support local decision-makers to make use of linked data to help families locally.

The interview will take place using Zoom or via telephone and can be arranged at time convenient for you. Interviews usually last around 30 minutes but can be adapted to suit your availability. All the information you need to take part will be provided to you in advance of the interview and you can contact Hollie to ask any questions you may have about the project. Your participation will remain anonymous, and any published data will not include identifiable information about participants. Participation in this study is entirely your choice. Please find attached the Participant Information Sheet.

You are eligible to take part in this research if you make decisions about perinatal or early life health and are located within either Bradford, Leeds, Wakefield, or Doncaster, or involved in the administration of a local electronic birth cohort study part of the Born and Bred in (BaBi) Network. If you are eligible and would like to take part in this research, contact Hollie Henderson by *insert contact details.

Please feel free to circulate this invite to colleagues you think would be interested in taking part. We would very much welcome your contribution to this important and timely research.

D2. Qualitative interview participant information sheet (Final version V1.1 19/10/2021)

We would like to invite you to take part in our research study, which is part of a PhD project funded by the White Rose Consortium. Joining the study is entirely your choice. Before you decide whether you would like to take part, please read the following information. Please feel free to talk to others about the study and to contact Hollie Henderson if you have any questions.

What is the purpose of this study?

An exciting new family of projects, known as Born and Bred in (BaBi), is being developed across the region, to help us learn more about how families, in our area and beyond, can live healthier, happier lives. Each local area has set up their own project, which gains consent from women during their pregnancy to link information routinely collected as part of everyday services about themselves and their baby. BaBi builds on the success of Born in Bradford 4 All (BiB4All), where Bradford is supporting other local sites to develop their projects. The term *linked data* refers to the linking of individual information from more than one source for the purposes of research.

The aim of this study is to explore how we can best engage with and support those working in the field of early life health to make use of this linked data to inform decision-making. This is important as we want to make best use of research produced as part of the BiB4All and BaBi project to improve services for families locally.

Who is doing the study?

Hollie Henderson, a PhD student located in the Department of Health Sciences at the University of York will be carrying out this study, in partnership with the Bradford Institute of Health Research. This will be under the supervision of Professor Kate Pickett, Professor James Wilsdon and Sally Bridges. This research is funded by the White Rose Consortium and is part of the National Institute for Health and Care Research Yorkshire and Humber Applied Research Collaboration.

Why have I been asked to participate?

You are invited to take part because you are a decision-maker working in the field of maternity, perinatal and early life health, in a local area which is part of the BaBi project

or are involved in the administration of a local BaBi project. We are inviting approximately 20 participants like you to take part. Your responses will be used to design a toolkit to support the BaBi research teams to engage in activities that allow their research to be relevant to decision-makers.

Do I have to take part?

It is entirely up to you to decide whether you would like to take part. If you have any questions, you are welcome to talk to Hollie Henderson, contact details will be provided at the end. If you do decide to take part, you will be given this information sheet to keep and be asked to sign an online consent form. We hope that as many decision-makers as possible take part as this will help ensure a range of views are included.

What will be involved if I take part in this study?

If you do decide to take part, contact Hollie Henderson to arrange a convenient time for you to attend an online interview held over Zoom. The interview will take no longer than 60 minutes and, with your permission, will be recorded via Zoom and on a Dictaphone. This will enable the discussion to be typed up in writing and ensure your views are fully captured. This recording will only be accessible to the research team supporting this project and will be destroyed after the recording has been typed up in writing. You can also take part via a telephone call if this is preferred. During the interview you can ask any questions about the study. You will be asked about your perspectives towards linked data from a consented cohort such as BaBi, its uses for early years decision-making and ways that could help support you to use research produced using this linked data in your decision-making. We also want to hear your views about how we can best engage you in the BaBi project so that you get the most out of the research that is produced. Once the data are analysed for the project, we may contact you again to ascertain your views on the findings and outcomes of the study. This will occur within six months of the interview.

What are the advantages or benefits of taking part?

There are no direct benefits to you taking part in this study. Any benefits will be indirect as a result of helping to inform the development of a toolkit that will engage with and support early years decision-makers to make use of linked data research to help families locally.

What are the disadvantages or risks of taking part?

We do not anticipate that there will be any disadvantages of taking part in this study. Your employment will not be affected by your decision to take part or not in the study.

In the reporting of the results, individuals will not be identified, but job roles will be described. Therefore, there are some instances where identification of participants is possible (e.g., where only one person fulfils that professional role). Further precautions will be taken to reduce the risk of identifying individuals such as no quotations directly attributable to a participant will be included and generic job descriptions will be used.

How will the information and personal data I give be used?

We will need to use information from you for this research project. To safeguard your rights, we will use the minimum personally identifiable information possible.

The information that we collect for this research will include:

- The information from our interview with you, which will be typed up in writing by the research team.
- The information from your completed electronic consent form.
- Your name and contact details. These details are collected for the purpose of arranging and carrying out the interviews and to contact you with the results, should you wish to be kept informed.

We will use your personal information only to do the research. Contact details will be destroyed following the interviews, unless you wish to be kept informed of the study findings, then your details will be destroyed following the completion of the final report. Other authorised individuals may check your records to make sure that the research is being done properly.

Your name and contact details will be kept separate from the other information that we obtain from you for this research. Your research records will contain a pseudonym (false name) instead, so you are not directly identifiable when we use your data in our research.

Your name, contact details and research records will be stored electronically on secure University of York approved storage. Access to this information will be restricted to authorised persons only. We will keep all information about you safe and secure.

For the purposes of obtaining consent to take part in this study, your data will be stored online via Qualtrics and on the secure University of York server. Please use the following link to view the privacy statement: https://www.qualtrics.com/privacy-statement/.

Once we have finished the study, some of the data will be kept and stored securely for ten years. This will enable us to complete our publications and reports. After this time, it will be securely destroyed. We will write our reports in a way that no-one can work out that you took part in the study. The data may also be shared with other researchers to conduct further related research, which is subject to approval.

What are my choices about how my information is used?

The University of York is a publicly funded organisation that conducts research to improve health and healthcare services. In legal terms, we are using your information for this research as part of 'a task in the public interest'. The ability to change the data that we have collected, however, is limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate.

If you decide to take part and then later change your mind, you can withdraw from the study without giving a reason by contacting Hollie Henderson. You may request that your data is not used in the study up to 28 days after the interview. We will destroy your data upon this request unless you give your permission for the research team to retain this. After this time, data gathered may have started to be analysed and used in the study findings. Any information that has been used in the study findings cannot be withdrawn. This would not affect your legal rights.

Where can you find out more about how your information is used?

All data will be handled in accordance with the General Data Protection Regulation (GDPR) principles. To find out more about GDPR please use the links provided:
https://www.york.ac.uk/records-management/dp/
https://www.york.ac.uk/records-management/dp/guidance/gdprcompliantresearch/
https://www.york.ac.uk/records-management/dp/your-info/generalprivacynotice/

You can find out more about how we use your information:

- At https://www.york.ac.uk/healthsciences/research/trials/trials-gdpr/
- At https://www.york.ac.uk/healthsciences/research/trials/trials-gdpr/research-partcipants/

- by sending an email to Hollie Henderson
- by emailing the University of York's data protection officer on dataprotection@york.ac.uk

Will my taking part in this study be kept confidential?

If you decide to take part in the study, what you tell us will be kept confidential. No one outside the research team will know that you have taken part in the study.

We will write our reports in a way that no-one can work out that you took part in the study. Data collected for the study may be looked at by authorised persons who are organising the research. Data may also be looked at by other authorised people to check that the study is being carried out correctly. All have a duty of confidentiality to you as a research participant.

The only time we would break our duty of confidentiality is if we are worried that you – or someone else – was being, or was likely to be, harmed. If that happens, we will talk with you about it.

What will happen to the results of the study?

Once all the interviews have been typed up in writing, they will be analysed by the research team, and a summary report will be sent to you before being disseminated more widely amongst the BaBi Network and the Yorkshire and Humber Applied Research Collaboration. You will have the opportunity to comment on the results before they are shared more widely. The findings will then be written up for publication in peer reviewed journals and may be presented at scientific conferences, so that other researchers can learn from our findings. Data will be treated confidentially and any output resulting from this study will only report data that does not identify individual participants.

One way we can get the most benefit from this work is to make the study data available to researchers for related research at the end of this study. The study data may be reused by the research team or researchers in other institutions but will not be used or released in such a way that you could be identified. Applications to reuse the data once the project is completed will be reviewed by the Born in Bradford Executive Committee.

Who has reviewed and approved this study?

The study has been reviewed and approved by the University of York's Health Sciences Research Governance Committee reference: *HSRGC/2022/488/A.*

Who do I contact for more information about the study?

If you have any questions or would like more information about the research, please contact Hollie Henderson. You can do this by *insert contact details*.

Who do I contact in the event of a complaint?

If you would like to talk to someone about a complaint, you can contact Sally Bridges, the lead supervisor for this research, by email: Sally.Bridges@bthft.nhs.uk.

The University of York is the data controller for the information collected for this research, which means we are responsible for looking after your personal information and using it properly. If you are unhappy about the way your personal data has been handled, you have the right to complain to the University's Data Protection Officer at dataprotection@york.ac.uk; if you are still unsatisfied, you have a right to report concerns to the Information Commissioner's Office at www.ico.org.uk/concerns.

Thank you for taking the time to read this information sheet and considering whether to take part in this study.

D3. Qualitative interview supplementary information sheet (V1.0 20/10/2021)

Background information

In 2019, Bradford initiated an exciting new electronic birth cohort study known as Born in Bradford 4 All (BiB4All). It asks women during their pregnancy for permission to access and use information that is routinely collected about them and their child for research purposes. This allows for data such as health records, social care, and education to be linked for those individuals, building a clearer picture of families lives over time. Once linked, it is expected that the information will be used by researchers, in collaboration with local services and commissioners, to help improve health, care, and services through research and planning. The term *linked data* is used to describe the linking of individual information from more than one source for the purposes of research. Bradford is now supporting teams across the region to establish electronic birth cohorts in other local areas, including Doncaster, Wakefield, and Leeds, referred to as Born and Bred in (BaBi). This means they can be used locally as health intelligence tools to shape services and brought together as one to answer research questions of national relevance.

A series of two-hour online workshops have taken place across Bradford, Wakefield, and Doncaster to prioritise research questions to be answered using data from the electronic birth cohorts. Attendees of the workshops were from a variety of backgrounds including parents, midwives, health visitors, clinicians, service commissioners and researchers. Each workshop was split into two parts. The first part placed attendees into groups to discuss areas of child and maternal health they thought were important for linked data research. The second part placed attendees into new groups to prioritise the suggestions put forward by groups in the first part in order of urgency and importance.

Following the workshops, the prioritised research ideas were narrowed down into those that could be answered with the data currently available in the BiB4All study. An additional session then brought together relevant stakeholders to coproduce the final research question to be answered as part of this research project. The remaining research priorities fed into the work of the Early Life and Prevention Theme of the NIHR Yorkshire and Humber Applied Research Collaboration.

We are interested in how you can then make use of the linked data research in your decision-making and how you feel you can be supported to do so. We also want to know

how you would like to be engaged in the BiB4All and BaBi research projects in the future that would allow you to make use of them as a resource. Your insights will be invaluable in helping to develop a toolkit that will support you to make use of linked data research to improve services for families locally.

D4. Qualitative interview e-consent form (Final version V1.0 20/10/2021)

Please note: your responses to this form will at all times remain confidential. You are consenting to having this information collected and stored by submitting this form.

I confirm I have read and understood the participant information version 1.1 dated 19/10/2021 for the above study.

- Yes
- No

I have had the opportunity to consider the information, ask questions and discuss this study.

- Yes
- No

I have received satisfactory answers to all my questions.

- Yes
- No

I have received enough information about the study.

- Yes
- No

I understand that my participation is voluntary and that I am free to withdraw at any time until 28 days after the interview without giving any reason and without my legal rights being affected. I understand that should I withdraw, then the information collected will be erased, unless I give permission for the research team to retain this.

- Yes
- No

I give permission for the interview to be video recorded, which will be saved securely for purpose of review by the researcher.

- Yes
- No

I understand that the study data may be stored and used in relevant future research, including by researchers in other institutions, but the data will not be used or released in such a way that I could be identified.

- Yes
- No

I understand that any information I provide, including personal data, will be kept confidential, stored securely and only accessed by authorised individuals.

- Yes
- No

I understand that sections of my study documents may be looked at by responsible individuals' part of supervisory team or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my study documents.

- Yes
- No

I understand that if the researcher thinks that I or someone else might be at risk of harm, they may have to contact the relevant authorities.

- Yes
- No

I give permission for the research team to contact me within six months of the project to get my perspectives on the research findings and project outcomes.

- Yes
- No

I understand that any information I give may be included in published documents, but all information will be anonymised.

- Yes
- No

I would like to receive a summary of the results of the study.

- Yes
- No

I agree to take part in this study.

- Yes
- No

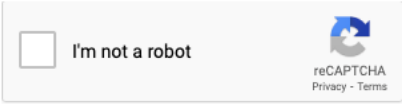
Enter your name

Date (DD/MM/YYYY)

Signature

SIGN HERE

I confirm I'm not a robot



We thank you for your time spent completing this consent form. Your response has been recorded.

Should you have any further questions, please contact Hollie Henderson by email: *contact details*

D5. Qualitative interview verbal consent form (Final version V1.0 20/10/2021)

	Question	Please initial this column to confirm verbal consent was given
1.	I confirm I have read and understood the participant information version 1.1 dated 19/10/2021 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2.	I understand that my participation is voluntary and that I am free to withdraw at any time without giving reason until 28 days after the interview. I understand that should I withdraw, then the information collected will be erased, unless I give permission for the research team to retain this.	
3.	I agree to this consent form and other data collected as part of this research study being kept at the University of York.	
4.	I understand that the study data may be stored and used in relevant future research, including by researchers in other institutions, but the data will not be used or released in a way that I could be identified.	
5.	I agree to the interviews being audio recorded and sections transcribed.	
6.	I understand that direct quotations may be used in publications, but no information will be released or printed that would identify me.	
7.	I understand and agree that any information I provide, including personal data, will be kept confidential and stored securely.	
8.	I give permission for the research team to contact me within six months of the project to get my perspectives on the research findings and project outcomes and to receive a summary of the findings.	
9.	I understand that sections of my study documents may be looked at by responsible individuals' part of supervisory team or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my study documents.	
10.	I agree to take part in the above study.	

_____ _____

Name of participant *(please print)* Date verbal consent given

_____ _____

Name of person taking consent Date

(please print)

D6. Interview topic guide

Research aims to explore:

- Perspectives towards linked data from an electronic birth cohort study and its use in early years decision-making
- Potential ways of engaging decision-makers in linked data research
- Views on the tools developed to support engagement of stakeholders in linked data research
- Perceived barriers and facilitators to using linked data from an electronic birth cohort study in early years decision-making

1. Introduction

- Introduce myself and my links to the BiB4All and BaBi project
- Explain:
    - Nature and purpose of the research, this includes details BaBi.
    - We are interested in how they could use this information in their decision-making focusing on using data from multiple data sources.
    - Background into what has been done so far in the project/reminder of what was discussed in the supplementary information sheet
    - Who the research is for?
    - The session will be recorded and the purpose for this
- Verbal consent to record and take part (if e-consent is not completed)
- Stress confidentiality
    - They may give examples from their own work but please be mindful of patient confidentiality and not to reveal any details that may identify a patient.
- Say what will happen with the research findings
- Ask if there are any questions at this stage

2. Background

- Job role, place of work and time spent in that role
- What area of perinatal or early life health they work in
- Decision-making experience in early life health

3. Experience of linked routine data

- What is their understanding and knowledge of linked routine data?
- Have they engaged with linked data research previously?

    - If not, what evidence do they use to inform their decision-making
- What are their views towards linked data evidence from a consented cohort?
- What do they think the benefits are of using this linked routine data in decision-making?
- What do they think the drawbacks are?
- How do they think intelligence derived from BiB4All/ BaBi could support their decision-making?

4. **Engaging decision-makers in linked data research**
    - How would they like to be involved in the linked data research process that would help them to use the evidence produced from BaBi/BiB4All?
        - Prompts include prioritisation of research questions, involvement in data analysis, involved in presenting the outputs?

5. **Views about current ways of engaging decision-makers in linked data research**
    - With reference to the information sheet, how do they view the process of prioritising research questions to be answered with linked data?
        - Remind them of the workshop format, if needed.
        - What support might they need to generate research priorities.

Remind participants that the workshop outputs will lead to the production of research evidence.

- Could they use this evidence in their decision-making?
    - If yes, how?
    - If no, why? What evidence would they use instead and why?

6. **Factors enabling or preventing the use of linked data in decision-making**
    - What would prevent from using linked data research from a consented birth cohort study?
        - Could linked data evidence be improved?
        - Do they need any additional skills or training to use the linked data research?
    - If they have used linked data previously, what has enabled them to use it?
    - What could researchers do to support them to use linked data research in their decision-making?

Summarising question

Overall, how do they view the use of linked data as evidence in early years decision-making?

Identifying additional participants

- Do they know any other decision-makers who might be interested in taking part in this study?
- Are they in touch with these individuals?
- Permission to follow up with them after the meeting to be put in touch with these individuals.

Next steps

- Thank the participant. Check whether they have any remaining questions about the research.
- Reassure them about confidentiality and anonymity.
- Inform them about the feedback email they will receive.

D7. Ethical considerations

Conducting research within an NHS setting

Research conducted within the NHS usually requires approval from the Health Research Authority (HRA). I sought this approval, however the HRA responded saying that it was not required. Thus, I approached the NHS Research and Development (R&D) teams, who support and facilitate high quality research within the NHS, at the NHS trusts where I was planning to carry out my research (Oxford University Hospitals NHS Foundation Trust, 2023). In the absence of HRA approval, the NHS R&D teams in two of the local areas part of the BaBi Network were apprehensive about my email invitation being shared as part of the BaBi steering group's regular email correspondence, as this involved NHS employees. To reassure the R&D teams, I forwarded the confirmation email from the HRA that approval was not needed and the accompanying reasons. I worked with them to ensure they were comfortable with the planned research, completed all relevant documentation, and undertook training including Good Clinical Practice training by the National Institute for Health and Care Research.

Consent

Email invitations included a Participant Information Sheet, which explained that participants had the right to withdraw from the study up to 28 days post interview, as after this point, it would be difficult to remove individual data from the analysis.

To ensure individuals fully understood the purpose of the study, the extent of their involvement and the risks and benefits of taking part, I presented details of the research study to the BaBi steering groups. For those recruited through other means such as through referrals of personal contacts, I provided more information on the study once they had contacted me with an interest in being involved.

Individuals were informed verbally that the interview was being recorded, for what purpose, and how and with whom this will be shared. A pop-up notice was displayed on the screen when the recording started so that the meeting participants were aware the recording had started.

Confidentiality and anonymity

A waiting room was enabled on the Zoom meeting to ensure that only the interviewer and participants were able to enter the online space, mitigating the risk of a non-participant entering the interview.

As set out in the Participant Information Sheet, the anonymised transcripts may be used by other researchers. I ensured names, names of services and other potentially identifiable information were replaced with generic descriptions so that the transcripts were still meaningful, whilst protecting confidentiality and anonymity.

Although I asked for participants' own perspectives, there was the potential for individuals working in the NHS to include examples that reveal patient information. Therefore, at the start of each interview, participants were reminded to respect patient confidentiality and not to disclose any information that could identify a patient. Their duty to patient confidentiality is set out in their employment and therefore it is their professional responsibility to prevent the disclosure of confidential information.

Finally, there was a risk that staff may feel obliged to take part. This was mitigated by emphasising that taking part was entirely voluntary, would have no bearing on their employment and that their participation is confidential.

Data management

I consulted with the University of York Health Sciences IT department to understand the most secure and safe method of storing and transferring the data collected.

In accordance with ethical guidance, contact details of participants were stored securely in a password protected file, to be used for the purposes of arranging the interviews. These details were kept until the project was completed and they had received a copy of the final report. If verbal consent was recorded during the interview, this was stored in a password protected folder on my University of York file store.

After each interview, the Zoom recording was downloaded to a University of York laptop where it was checked to see if it was complete. If the recording was intact and complete, it was retained and stored securely on University of York approved storage and the Dictaphone recording was erased. If the Zoom recording was corrupted or there was a failure to record, the Dictaphone recording was uploaded to the University of York approved storage and the Zoom recording erased. Interviews conducted over the phone were recorded on a Dictaphone. The recordings were used to transcribe the conversation and once the analysis process was completed, the recordings were deleted.

Interview recordings were transcribed using Otter.ai, which can only be viewed on my personal password protected account. It complies with data security and privacy policies: https://help.otter.ai/hc/en-us/articles/360048258953-Data-security-and-privacy-policies.

Transcripts were pseudonymised and stored securely on my University of York file store. They were uploaded to a qualitative data management software (NVIVO 12.1) for analysis, which complies with the relevant data privacy and handling laws. A copy of the transcription was saved to University of York approved storage, Google Drive, to allow access for my supervisory team, where required.

Following the completion of my PhD, pseudonymised transcripts will be stored securely for ten years on servers in the Bradford Teaching Hospitals NHS Foundation Trust Storage Area Network with physical accessed securely managed by the BTHFT IT department. These transcripts will be accessible for subsequent publications and secondary research, subject to approval by the Born in Bradford Executive Committee. This is appropriate as this PhD is in partnership with Born in Bradford and their BiB4All project. This was communicated to participants in the Participant Information Sheet. After this time, the transcripts will be securely destroyed.

Risks and benefits

This research was not believed to raise any potential for physical and/or psychological harm or distress to the participants or researchers. However, a review of the research processes took place at every stage so that any issues that arose could be acted upon promptly.

References

Abrams, K., Wang, Z., Sing, Y. and Galindo-Gonzalez, S. (2015) 'Data Richness Trade-offs Between Face-to-Face, Online Audiovisual, and Online Text-Only Focus Groups.' *Social Science Computer Review,* 3: 80- 96.

ADR UK. (2022a). *What is 'research-ready' data? A report of the online roundtable event held on 11 February 2022* [Online]. Available at: https://www.adruk.org/fileadmin/uploads/adruk/Documents/What_is_research_ready_data__A_roundtable_report_June_2022.pdf [Accessed 28th May 2023].

ADR UK. (2022b). *Written evidence submitted by Administrative Data Research UK (ADR UK) (DDA0031)* [Online]. Available at: https://committees.parliament.uk/writtenevidence/43433/pdf/ [Accessed 09th May 2022].

ADR UK. (2021a). *Birth cohort data linkage study* [Online]. ADR UK. Available at: https://www.adruk.org/our-mission/our-impact/article/birth-cohort-data-linkage-study-228/ [Accessed 20th May 2021].

ADR UK. (2021b). *ECHILD: Linking children's health and education data for England* [Online]. Available at: https://www.adruk.org/our-work/browse-all-projects/echild-linking-childrens-health-and-education-data-for-england-142/ [Accessed 17th June 2021].

ADR UK. (2021c). *About ADR* [Online]. Available at: http://ADRuk.org [Accessed 17th June 2021].

ADR UK. (2021d). *What is administrative data?* [Online]. ADR UK. Available at: https://www.adruk.org/our-mission/administrative-data/ [Accessed 21st May 2021].

ADR, UK. (2021e). *Our public engagement activities* [Online]. Available at: https://www.adruk.org/our-mission/our-public-engagement-activities/ [Accessed 3rd February 2022].

Akincigil, A., Munch, S. & Niemczyk, K. C. (2010). Predictors of Maternal Depression in the First Year Postpartum: Marital Status and Mediating Role of Relationship Quality. *Social Work in Health Care,* 49, 227-244.

Al-Saadi, H. (2014). Demystifying Ontology and Epistemology in Research Methods. Online]. Available at: https://www.researchgate.net/publication/260244813 [Accessed 24th January 2024].

Archibald, M. M., Ambagtsheer, R. C., Casey, M. G., & Lawless, M. (2019). Using Zoom Videoconferencing for Qualitative Data Collection: Perceptions and Experiences of Researchers and *Participants. International Journal of Qualitative Methods*, 18, 160940691987459. https://doi.org/10.1177/1609406919874596

Assarian, F., Moravveji, A., Ghaffarian, H., Eslamian, R. & Atoof, F. (2014). The association of postpartum maternal mental health with breastfeeding status of mothers: a case-control study. *Iranian Red Crescent Medical Journal,* 16**,** e14839.

Audrey, S., Brown, L., Campbell, R., Boyd, A. & Macleod, J. (2016). Young people's views about consenting to data linkage: findings from the PEARL qualitative study. *BioMed Central Medical Research Methodology,* 16**,** 34.

Barrett, E. S. *et al.,* (2014). Exposure to prenatal life events stress is associated with masculinized play behavior in girls. *Neurotoxicology (Park Forest South),* 41, pp.20–27. [Online]. Available at: doi:10.1016/j.neuro.2013.12.011. [Accessed 9th February 2024].

Bartley, M. (1992) *Authorities and Partisans: Debate on Unemployment and Health*. Edinburgh: Edinburgh University Press.

Bartley, M. (1988) *Unemployment & Health 1975–1987: A Case Study in the Relationship Between Research & Policy Debate*. PhD Thesis. Edinburgh: University of Edinburgh.

Battersby, C., Statnikov, Y., Santhakumaran, S., Gray, D., Modi, N. & Costeloe, K. (2018). The United Kingdom National Neonatal Research Database: A validation study. *PloS One,* 13**,** e0201815-e0201815.

Bauer, A., Parsonage, M., Knapp, M., Lemmi, V. & Adelaja, B. (2014). The costs of perinatal mental health problems. Personal Social Services Research Unit, Centre for Mental Health.

Baum, C. (2003). Does Early Maternal Employment Harm Child Development? An Analysis of the Potential Benefits of Leave Taking. *Journal of Labor Economics,* 21**,** 381-408.

Better Start Child Health and Development Research Group. (2014). *The South Australian Early Childhood Health and Development Project*
[Online]. The University of Adelaide, School of Population Health. Available at: https://health.adelaide.edu.au/betterstart/ua/media/100/data-linkage-project-description.pdf [Accessed 7th August 2023].

Bhaskar, R., (1975). *A realist theory of science*. Leeds: Leeds Books Ltd.

Bhaskar, R. (1979). *The possibility of naturalism: A philosophical critique of the contemporary human sciences*. Atlantic Highlands, NJ: Humanities Press.

Black, N. (2001). Evidence based policy: proceed with care. *British Medical Journal,* 323**,** 275-9. doi: 10.1136/bmj.323.7307.275.

Black, M. M., Baqui, A. H., Zaman, K., El Arifeen, S. & Black, R. E. (2009). Maternal depressive symptoms and infant growth in rural Bangladesh. *American Journal of Clinical Nutrition,* 89**,** 951s-957s.

Blau, F. & Grossberg, A. (1992). Maternal Labor Supply and Children's Cognitive Development. *The Review of Economics and Statistics,* 74**,** 474-81.

Blum, S., and Pattyn, V. (2022). How are evidence and policy conceptualised, and how do they connect? A qualitative systematic review of public policy literature. *Evidence & Policy* 18, 3, 563-582.

Booth, A., Papaioannou, D. & Sutton, A. (2016). *Systematic approaches to a successful literature review*, SAGE.

Born in Bradford (2023). *BiB4All*. [Online]. Born in Bradford. Available at: https://borninbradford.nhs.uk/what-we-do/cohort-studies/born-in-bradford/ [Accessed 15th August 2023].

Born in Bradford (2021). *Born in Bradford About Us.* [Online]. Born in Bradford. Available at: https://borninbradford.nhs.uk/about-us/ [Accessed 15th January 2021].

Born in Bradford. (2020). *How we use your Information – Born in Bradford*. [Online]. Available at: https://borninbradford.nhs.uk/what-we-do/how-we-use-your- information/ [Accessed 19 November 2020].

Born In Bradford. (2019). Born in Bradford: 10 Key Findings for Policy and Practice. [Online]. *Born in Bradford* Available at: https://borninbradford.nhs.uk/wp-content/uploads/Key- Findings_FINAL_Designed.pdf [Accessed 15th January 2021].

Born in Bradford (2018). Born in Bradford (BiB): A data linkage cohort study of babies born in Bradford and their mothers: Information for pregnant women. [Online]. Born in Bradford Available at: https://borninbradford.nhs.uk/wp-content/uploads/Born-in-Bradford-A3-Participant-booklet-V-4.1-04.10.18.pdf [Accessed 15th January 2021].

Borra, C., Iacovou, M. & Sevilla, A. (2015). New Evidence on Breastfeeding and Postpartum Depression: The Importance of Understanding Women's Intentions. *Maternal and Child Health Journal,* 19**,** 897-907.

Boswell, C. & Smith, K. (2017). Rethinking policy 'impact': four models of research-policy relations. *Palgrave Communications,* 3**,** 44.

Boswell, C., Smith, K., Davies, C., (2022). *Rethinking Policy Impact: Promoting Ethical and Effective Policy Engagement in the Higher Education Sector* [*online*]: The Royal Society Edinburgh [Online]. Available at: https://rse.org.uk/wp-content/uploads/2022/10/RSE-Rethinking-Policy-Impact-report.pdf [Accessed 7th August 2023].

Bowe, A. K., Hourihane, J., Staines, A. & Murray, D. M. (2022). The predictive value of the ages and stages questionnaire in late infancy for low average cognitive ability at age 5. *Acta Paediatrica,* 111**,** 1194-1200.

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S. & Smith, G. D. (2013). Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology,* 42**,** 111-127.

Boyle, D. & Emery, J. (2017). Data linkage. *Australian Journal for General Practitioners* 46 615-619.

Bradford Institute for Health Research. (2023). *Connected Yorkshire* [Online]. Available at: https://www.bradfordresearch.nhs.uk/our-research-teams/connected-bradford/ [Accessed 8th August 2023].

Bradford Institute for Health Research. (2020) Born in Bradford: A data linkage cohort study of babies born in Bradford and their mothers: protocol May 2020 Version 4.0. Unpublished.

Bradford Teaching Hospitals NHS Foundation Trust. (2023). *Community Midwifery* [Online]. Available at: https://www.bradfordhospitals.nhs.uk/community-midwifery/ [Accessed 23rd August 2023].

Bradford Teaching Hospitals NHS Foundation Trust. (2021a). *BaBi Bradford (BiB4All)* [Online]. Available at: https://www.babinetwork.co.uk/babi-sites/babi-bradford [Accessed 23rd June 2023].

Bradford Teaching Hospitals NHS Foundation Trust. (2021b). *Born and Bred in (BaBi)* [Online]. Available at: https://www.babinetwork.co.uk [Accessed 23rd June 2023].

Braun, V. and V. Clarke (2023). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing researcher. *International Journal of Transgender Health* 24(1) 1-6.

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.

Braun, V. & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health,* 13**,** 201-216.

Braun, V. & Clarke, V. & (2013). *Successful Qualitative Research: A Practical Guide for Beginners*. SAGE.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brentani, A. & Fink, G. (2016). Maternal depression and child development: Evidence from São Paulo's Western Region Cohort Study. *Revista Associacao Medica Brasileira (1992),* 62**,** 524-529.

Briggs-Gowan, A. J., Carter, M. S. and Schwab-Stone, M. (1996). Discrepancies among mother, child, and teacher reports: Examining the contributions of maternal depression and anxiety. *Journal of Abnormal Child Psychology*, 24 (6), pp.749–765. [Online]. Available at: doi:10.1007/BF01664738.

British Health Foundation. (2012). *Using big data to improve heart health* [Online]. Available at: https://www.bhf.org.uk/research-projects/mrc-ehealth-centresfor-health-informatics-research-in-the-uk-linkage-and-analysis-of-ehealth-data-across-the-uk [Accessed 09/05/2022 2022].

Brookes Publishing Co Inc. (2022). *ASQ-3* [Online]. Available: https://agesandstages.com/products-pricing/asq3/ [Accessed 26th May 2022].

Bryant, M. & Bridges, S. (2021). *Born and Bred in Protocol*. Unpublished

Bryar, R., Anto-Awuakye, S., Christie, J., Davis, C., & Plumb, K. (2013). Using the Delphi Approach to Identify Priority Areas for Health Visiting Practice in an Area of Deprivation. *Nursing Research and Practice*, *2013* (page 21), 1-9. https://doi.org/10.1155/2013/780315

Buxton, M., & Hanney, S. (1996). How can payback from health services research be assessed?. *Journal of Health Services Research & Policy*, *1*(1), 35–43.

Cabinet Office. (1999) *Modernising Government* (White Paper). London: The Stationary Office.

Cairney P. (2013). *Policy Concepts in 1000 Words: The Policy Cycle and its Stages* [Online]. Paul Cairney: Politics & Public Policy. Available at: https://paulcairney.wordpress.com/2013/11/11/policy-concepts-in-1000-words-the-policy-cycle-and-its-stages/ [Accessed 29th May 2023 2023].

Cairney, P. & Kwiatkowski, R. (2017). How to communicate effectively with policymakers: combine insights from psychology and policy studies. *Palgrave Communications,* 3**,** 37.

Cairney, P. & Oliver, K. (2018). How Should Academics Engage in Policymaking to Achieve Impact? *Political Studies Review,* 18**,** 228-244.

CALLS-HUB. (2021a). *About* [Online]. Available: https://calls.ac.uk/about/ [Accessed 12th June 2021].

CALLS-HUB. (2021b). *Impact Case Studies* [Online]. Available at: https://calls.ac.uk/outputs/impact-case-studies/ [Accessed 12th June 2021].

Cao, X. *et al.,* (2014). Prenatal maternal stress affects motor function in 5½-year-old children: Project Ice Storm. *Developmental psychobiology*, 56 (1), pp.117–125. [Online]. Available at: doi:10.1002/dev.21085[Accessed 9th February 2024].

Caplan, N. (1979) The two-communities theory and knowledge utilization. *American Behavioral Scientist* 22(3):459-470.

Carlsen, B., & Glenton, C. (2011). What about N? A methodological study of sample-size reporting in focus group studies. *BioMed Central Medical Research Methodology*, *11*(1), 1-10.

Carson, L. E., Azmi, B., Jewell, A., Taylor, C. L., Flynn, A., Gill, C., Broadbent, M., Howard, L., Stewart, R. & Poston, L. (2020). Cohort profile: the eLIXIR Partnership—a maternity-child data linkage for life course research in South London, UK. *British Medical Journal open,* 10**,** e039583-e039583.

Centre For Longitudinal Studies. (2021). *Millennium Cohort Study* [Online]. Available at: https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/ [Accessed 12th June 2021].

Centre For Longitudinal Studies. (2011). *Parents' marital status does not significantly improve children's outcomes* [Online]. Available: https://cls.ucl.ac.uk/parents-marital-status-does-not-significantly-improve-childrens-outcomes/ [Accessed 2nd August 2022].

Charkaluk, M.-L., Rousseau, J., Calderon, J., Bernard, J. Y., Forhan, A., Heude, B., & Kaminski, M. (2017). Ages and Stages Questionnaire at 3 Years for Predicting IQ at 5–6 Years. *Pediatrics,* 139**,** e20162798.

Christen P., and Schnell R. (2023). Thirty-three myths and misconceptions about population data: From data capture and processing to linkage. *International Journal of Population Data Science* **8**:1-14. doi: 10.23889/ijpds.v8i1.2115

City of Bradford Metropolitan District Council. (2023). *Poverty and deprivation: Deprivation and poverty in Bradford District* [Online]. Available at: https://ubd.bradford.gov.uk/about-us/poverty-in-bradford-district/2023].

CLOSER. (2018a). *Maximising research impact* [Online]. Available at: https://www.closer.ac.uk/maximising-research-impact/ [Accessed 12th June 2021].

CLOSER. (2018b). *Hertfordshire Cohort Study* [Online]. Available at: https://www.closer.ac.uk/study/hertfordshire-cohort-study/ [Accessed 12 June 2021].

CLOSER. (2018c). *Wirral Child Health and Development Study* [Online]. Available at: https://www.closer.ac.uk/study/wirral-child-health-and-development-study/ [Accessed 15th June 2021].

CLOSER. (2018d). *Understanding Society: The UK Household Longitudinal Study* [Online]. Available at: https://www.closer.ac.uk/study/understanding-society/ [Accessed 15th June 2021].

CLOSER. (2018e). *About CLOSER* [Online]. Available: at https://www.closer.ac.uk/home/what-we-do/ [Accessed 15th June 2021].

CLOSER. (2018f). *Explore the Studies* [Online]. Available at: https://www.closer.ac.uk/explore-the-studies/ [Accessed 15th June 2021].

Clowes, M., Sutton, A. (2021): "It's A Grey Area": searching the grey literature on how local governments use real-world data. The University of Sheffield. Report. https://doi.org/10.15131/shef.data.16644916.v1

Cockerham, W. C., Dingwall, R. and Quah, S (2014). *Socioeconomic Status, Definition*. John Wiley. [Online]. Available at: doi:10.1002/9781118410868.

Cogill, S. R., Caplan, H. L., Alexandra, H., Robson, K. M. & Kumar, R. (1986). Impact of maternal postnatal depression on cognitive development of young children. *British Medical Journal (Clinical research ed.),* 292**,** 1165-1167.

Cohen, M., March, J. and Olsen, J. (1972) A Garbage can model of organizational choice. *Administrative Science Quarterly* 17:1-25.

Collins Dictionary (2023a). *Definition of 'policy'.* [Online]. Collins Dictionary. Available at: https://www.collinsdictionary.com/dictionary/english/policy [Accessed 26 May 2023].

Collins Dictionary (2023b). *Definition of 'local authority'.* [Online]. Collins Dictionary. Available at: https://www.collinsdictionary.com/dictionary/english/local-authority [Accessed 26 May 2023].

Contandriopoulos, D., Lemire, M., Denis, J.-L. and Tremblay, É. (2010) Knowledge exchange processes in organizations and policy arenas: a narrative systematic review of the literature. *Milbank Quarterly* 88(4): 444–483.

Concannon, T. W., Fuster, M., Saunders, T., Patel, K., Wong, J. B., Leslie, L. K. & Lau, J. (2014). A Systematic Review of Stakeholder Engagement in Comparative Effectiveness and Patient-Centered Outcomes Research. *Journal of General Internal Medicine,* 29**,** 1692-1701.

CPRD. (2021). *CPRD linked data* [Online]. Available at: https://cprd.com/linked-data [Accessed 17th June 2021].

Cresswell, K., Morrison, Z., Sheikh, A. & Kalra, D. (2012). "There Are Too Many, but Never Enough": Qualitative Case Study Investigating Routine Coding of Clinical Information in Depression. *PloS One,* 7**,** e43831.

Crowe, S., et al. (2011). The case study approach. *BioMed Central Medical Research Methodology* 11(1) 100.

Daas, P. J., Ossen, S. J., Tennekes, M. & Burger, J. (2012) Evaluation and visualisation of the quality of administrative sources used for statistics. *Paper for the European Conference on Quality in Official Statistics, 2012*.

Dachew, B. A., Tessema, G. A. and Alati, R. (2023). Association between obstetric mode of delivery and emotional and behavioural problems in children and adolescents: the children of the 90s health study. *Social Psychiatry and Psychiatric Epidemiology,* 58 (6), pp.949–960. [Online]. Available at: doi:10.1007/s00127-022-02374-z.

Data Linkage Scotland. (2016). *Data Linkage Scotland: Information and case studies about data linkage research and services* [Online]. Available at: https://www.ubdc.ac.uk/media/1440/data-linkage-scotland-showcase-2016-brochure.pdf [Accessed 17th June 2021].

Data Protection Act 2018, c. 12. London: The Stationery Office.

Davies, H. T. O., Nutley, S. and Smith, P. C. (eds) (2000) *What Works? Evidence- Based Policy and Practice in Public Services*. Bristol: The Policy Press.

Davis-Kean, P., Chambers, R., Davidson, L., Kleinert, C., Ren, Q. & Tang, S. (2017). *Longitudinal Studies Strategic Review 2017 Report to the Economic and Social Research Council*. [Online]. Available at: https://www.ukri.org/publications/longitudinal-studies-strategic-review/ [Accessed 26th January 2024].

Davis, K., Sudlow, C. & Hotopf, M. (2016). Can mental health diagnoses in administrative data be used for research? A systematic review of the accuracy of routinely collected diagnoses. *BioMed Central Psychiatry,* 16.

De Man, Y., Wieland-Jorna, Y., Torensma, B., De Wit, K., Francke, A. L., Oosterveld-Vlug, M. G. & Verheij, R. A. (2023). Opt-In and Opt-Out Consent Procedures for the Reuse of Routinely Recorded Health Data in Scientific Research and Their Consequences for Consent Rate and Consent Bias: Systematic Review. *Journal of Medical Internet Research,* 25**,** e42131.

Deeny, S. R. & Steventon, A. (2015). Making sense of the shadows: priorities for creating a learning healthcare system based on routinely collected data. *British Medical Journal Quality & Safety,* 24**,** 505.

Dekel, S., Ein-Dor, T., Berman, Z., Barsoumian, I. S., Agarwal, S. & Pitman, R. K. (2019). Delivery mode is associated with maternal mental health following childbirth. *Archives Women's Mental Health,* 22**,** 817-824.

Denegri, S., et al., (2015). *Going the Extra Mile - Improving the Nation's health and wellbeing through public involvement in research*. [Online]. NIHR. Available at: https://www.nihr.ac.uk/documents/about-us/our-contribution-to-research/how-we-involve-patients-carers-and-the-public/Going-the-Extra-Mile.pdf. [Accessed 26th January 2024]

Department for Digital, Culture, Media & Sport & Department for Science, Innovation & Technology. (2020). *National Data Strategy* [Online]. Available at: https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy#data-3-4 [Accessed 30th May 2023].

Department of Health. (2022). *Factsheet: Developing a public health outcome measure for children aged 2 – 21⁄2 using ASQ-3TM* [Online]. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550487/ASQ-3_September_2016.pdf [Accessed 26th May 2022].

Department of Health. (2012). New focus for Public Health – the Health and Social Care Act 2012 (factsheet). [Online]. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/138263/B4.- Factsheet-New-focus-for-public-health-250412.pdf [Accessed 26th January 2015].

Department of Health. (2006) Best Research for Best Health. London

Department of Health and Social Care. (2022). *Data saves lives: reshaping health and social care with data* [Online]. Available at: https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data [Accessed 30th May 2023].

Department for Health and Social Care. (2021a). *Integration and innovation: working together to improve health and social care for all* [Online]. Available at: https://www.gov.uk/government/publications/working-together-to-improve-health-and-social-care-for-all/integration-and-innovation-working-together-to-improve-health-and-social-care-for-all-html-version [Accessed 30th May 2023].

Department of Health and Social Care. (2021b). The best start for life: a vision for the 1001 critical days. Available at: https://www.gov.uk/government/publications/the-best-start-for-life-a-vision-for-the-1001-critical-days [Accessed 26th May 2023].

Dias, C. C. & Figueiredo, B. (2015). Breastfeeding and depression: a systematic review of the literature. *Journal Affective Disorders,* 171**,** 142-54.

Dickerson, J., Bird, P. K., Mceachan, R. R. C., Pickett, K. E., Waiblinger, D., Uphoff, E., Mason, D., Bryant, M., Bywater, T., Bowyer-Crane, C., Sahota, P., Small, N., Howell, M., Thornton, G., Astin, M., Lawlor, D. A. & Wright, J. (2016). Born in Bradford's Better Start: an experimental birth cohort study to evaluate the impact of early life interventions. *BioMed Central Public Health,* 16**,** 711.

Digital Economy Act. (2017). Research Codes of Practice and Accreditation Criteria. Digital Economy Act 2017 part 5: Codes of Practice. Updated 2020 February.

DiPietro, J. A., Novak, M. F., Costigan, K. A., Atella, L. D. & Reusing, S. P. (2006). Maternal psychological distress during pregnancy in relation to child development at age two. *Child Development,* 77**,** 573-87.

Dobbins, M., Ciliska, D., Cockerill, R., Barnsley, J. & Dicenso, A. (2002). A Framework for the Dissemination and Utilization of Research for Health-Care Policy and Practice. *Worldviews on Evidence-based Nursing presents the archives of Online Journal of Knowledge Synthesis for Nursing,* E9**,** 149-160.

Dobbins, M., Jack, S., Thomas, H. & Kothari, A. (2007). Public health decision-makers' informational needs and preferences for receiving research evidence. *Worldviews Evidence-Based Nursing,* 4**,** 156-63.

Dobbins, M., Robeson, P., Ciliska, D., Hanna, S., Cameron, R., O'mara, L., Decorby, K. & Mercer, S. (2009). A description of a knowledge broker role implemented as part of a randomized controlled trial evaluating three knowledge translation strategies. *Implementation Science,* 4**,** 23.

Donmoyer, R. (2009). Generlizability and the Single-Case Study. In Gomm, R., Hammersley, M. & Foster, P. (eds.) *Case Study Method.* London: Sage.

Donnison, D. (1972) Research for policy. *Minerva.* 10(4): 519–536.

Dwyer, S. C. and J. L. Buckle (2009). The Space Between: On Being an Insider-Outsider in Qualitative Research. *International Journal of Qualitative Methods* 8(1) 54-63.

Early Intervention Foundation. (2021) *What early intervention is about: Early Years*. [Online]. Available at: https://www.eif.org.uk/what-its-about/early-years [Accessed 15th January 2022].

eCRUSADer. (2022). *Sharing Information. Improving Understanding* [Online]. Available at: https://ecrusad.co.uk [Accessed 16th August 2023].

Eichler, J., Schmidt, R., Poulain, T., Hiemisch, A., Kiess, W. & Hilbert, A. (2021). Maternal depressive symptoms and stress during pregnancy as predictors of gestational age at birth and standardized body mass index from birth up to 2 years of age. *BioMed Central Pregnancy Childbirth,* 21**,** 635.

Eisenstadt, S. N. (ed.) (1968) *Max Weber – On Charisma and Institution Building*. Chicago and London: The University of Chicago Press.

Ertel, K. A., Koenen, K. C., Rich-Edwards, J. W. & Gillman, M. W. (2010). Maternal Depressive Symptoms Not Associated with Reduced Height in Young Children in a US Prospective Cohort Study. *PloS One,* 5**,** e13656.

ESRC. (2018). *Longitudinal Studies Strategic Review: Annexes* [Online]. Available at: https://esrc.ukri.org/files/news-events-and-publications/publications/annexes-to-the-longitudinal-studies-review-2017/ [Accessed 17th June 2021].

Evans, K., Spiby, H. & Morrell, J. C. (2020). Non-pharmacological interventions to reduce the symptoms of mild to moderate anxiety in pregnant women. A systematic review and narrative synthesis of women's views on the acceptability of and satisfaction with interventions. *Archives of Women's Mental Health,* 23**,** 11-28.

Fadlallah, R., Daher, N., El-Harakeh, A., Hammam, R., Brax, H., Bou Karroum, L., Lopes, L. C., Arnous, G., Kassamany, I., Baltayan, S., Harb, A., Lotfi, T., El-Jardali, F. & Akl, E. A. (2022). Approaches to prioritising primary health research: a scoping review. *British Medical Journal Global Health,* 7**,** e007465.

Falster, K., Hanly, M., Banks, E., Lynch, J., Chambers, G., Brownell, M., Eades, S. & Jorm, L. (2018). Maternal age and offspring developmental vulnerability at age five: A population-based cohort study of Australian children. *PLoS Medicine,* 15**,** e1002558.

Family Nurse Partnership. (2022). *Enabling parents to give their children the best start in life* [Online]. Available at: https://fnp.nhs.uk [Accessed 15th February 2022].

Family Nurse Partnership National Unit & Dartington Service Design Lab. (2020). *FNP ADAPT: Using evidence, pragmatism and collaboration to change the FNP programme in England* [Online]. London, UK. Available at: https://www.fnp.nhs.uk/fnp-development/adapt [Accessed 21st November 2022].

Felfe, C. & Hsin, A. (2012). Maternal work conditions and child development. *Economics of Education Review,* 31**,** 1037-1057.

Fink, A, Kosecoff, J, Chassin, M, & Brook, R H. (1984). Consensus methods: Characteristics and guidelines for use. *American Journal of Public Health* (1971), 74(9), 979-983.

Fletcher, A. J. (2017). Applying critical realism in qualitative research: methodology meets method. *International Journal of Social Research Methodology,* 20(2)**,** 181- 194.

Flynn, A. J., *et al.,* (2018). The Knowledge Object Reference Ontology (KORO): A formalism to support management and sharing of computable biomedical knowledge for learning health systems. *Learning Health Systems,* 2(2) e10054.

Forbes, C., et al. (2022). Prioritising child health and maternity evidence-based interventions or service models: a stakeholder-driven process. *BioMed Central Health Services Research* 22(1) 764.

Ford, E., Campion, A., Chamles, D. A., Habash-Bailey, H. & Cooper, M. (2016). "You don't immediately stick a label on them": a qualitative study of influences on general practitioners' recording of anxiety disorders. *British Medical Journal Open,* 6**,** e010746.

Franklin, M., Howdon, D., Mason, S., Stone, T. & Jones, M. (2022). *Unlocking Real-World Data Offers Real Benefits to Public Health* [Online]. Social Sciences Space: Social Sciences Space. Available at: https://www.socialsciencespace.com/2022/01/unlocking-real-world-data-offers-real-benefits-to-public-health/?utm_source=twitter&utm_medium=SAGE_social&utm_content=socscispace&utm_term=704bc318-475a-4020-9f1d-d1347459a3ee [Accessed 26th May 2023].

Giallo, R., Woolhouse, H., Gartland, D., Hiscock, H. & Brown, S. (2015). The emotional-behavioural functioning of children exposed to maternal depressive symptoms across pregnancy and early childhood: a prospective Australian pregnancy cohort study. *European Child & Adolescent Psychiatry,* 24**,** 1233-1244.

Glaser, B., Strauss, A. (1999). *The discovery of grounded theory: Strategies for qualitative research.* New York: Aldine de Gruyter.

Glass, H.C., Costarino, A.T., Stayer, S.A., Brett, C., Cladis, F., Davis, P.J. (2015). Outcomes for extremely premature infants. *Anesthesia and Analgesia,* 120(6)**,** 1337-51.

Gleason, J. L., Gilman, S. E., Sundaram, R., Yeung, E., Putnick, D. L., Vafai, Y., Saha, A. & Grantz, K. L. (2021). Gestational age at term delivery and children's neurocognitive development. *International Journal of Epidemiology,* 50**,** 1814-1823.

Godlee. F. (2016) What can we salvage from care.data? *British Medical Journal*, 354. doi:10.1136/bmj.i3907

Goisis, A., Özcan, B. & Myrskylä, M. (2017). Decline in the negative association between low birth weight and cognitive ability. *Proceedings of the National Academy of Sciences,* 114**,** 84-88.

Goldacre, B. and Morely, J. (2022) Better, broader, safer: using health data for research and analysis. A review commissioned by the Secretary of state for health and social care. Department of Health and Social Care. [Online]. Available at: https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis [Accessed 26th May 2023].

Gomm, R., Hammersley, M. and Foster, P. (2009). Introduction. In Gomm, R., Hammersley, M. & Foster, P. (eds.) *Case Study Method,* London: Sage.

Goodman, J. H. (2004). Paternal postpartum depression, its relationship to maternal postpartum depression, and implications for family health. *Journal of Advanced Nursing,* 45**,** 26-35.

Government of Western Australia. (2023). *QA Health Data Linkage Strategy 2022 - 2024* [*Online*]. Available at: https://www.health.wa.gov.au/~/media/Corp/Documents/Reports-and-publications/WA-Health-Data-linkage-Strategy/WA-Health-Data-Linkage-Strategy-2022-2024.pdf [Accessed 8th September 2023].

Government of Western Australia Department of Health. (2023). *Data Linkage Services Western Australia* [*online*]. Available at: https://www.datalinkageservices.health.wa.gov.au [Accessed 17th September 2023].

Grant, M. J. & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal,* 26**,** 91-108.

Green, E., Ritchie, F., Mytton, J., Webber, D. J., Deave, T., Montgomery, A., Woolfrey, L., ul-Baset, K., & Chowdhury, S. (2015). *Enabling data linkage to maximise the value of public health research data*. https://uwe-repository.worktribe.com/output/836824

Greenhalgh, T., et al. (2016). Research impact: a narrative review. *BioMed Central Medicine,* 14(1) 78.

Growing Up in Scotland. (2021). *About GUS* [Online]. Available at: https://growingupinscotland.org.uk/about-gus/ [Accessed 8th June 2021].

Growing Up in Scotland (2012). *'Growing up in Scotland' Assessing contribution and impact June 2012.* [Online]. Available at: https://growingupinscotland.org.uk/wp-content/uploads/2018/08/GUS- ImpactReportwithAppendicesJune2012.pdf [Accessed 8th June 2021].

Growing Up in Scotland. (2010). *Using the findings from the Growing Up in Scotland study - a guide for Local Authorities*, Centre for Research on Families and Relationships, The University of Edinburgh.

Gualdi-Russo, E., et al. (2014). Health, growth and psychosocial adaptation of immigrant children. *European Journal of Public Health*,24(suppl_1) 16-25.

Hack, M., Klein, N. K. & Taylor, H. G. (1995). Long-term developmental outcomes of low birth weight infants. *Future Child,* 5**,** 176-96.

Ham, C., Hunter, D. J. and Robinson, R. (1995) 'Evidence based policymaking', *British Medical Journal*. British Medical Journal Publishing Group, p. 71. doi: 10.1136/bmj.310.6972.71

Hamdan, A. & Tamim, H. (2012). The relationship between postpartum depression and breastfeeding. *International Journal of Psychiatry Medicine,* 43**,** 243-59.

Hardicre, N. (2022) BaBi Leeds Workshop: Summary Report. Unpublished.

Harron, K. (2016). *Better Knowledge Better Society: An Introduction to Data Linkage* [Online]. Administrative Data Research Network. Available at: https://eprints.ncrm.ac.uk/id/eprint/4282/1/Guide%20to%20data%20linkage.pdf [Accessed 24 March 2021].

Harron, K., Dibben, C., Boyd, J., Hjern, A., Azimaee, M., Barreto, M. L. & Goldstein, H. (2017). Challenges in administrative data linkage for research. *Big Data & Society,* 4**,** 2053951717745678.

Harron, K., Gamble, C. & Gilbert, R. (2014). E-health data to support and enhance randomised controlled trials in the United Kingdom. *Clinical Trials,* 12**,** 180-182.

Harron, K., Wade, A., Muller-Pebody, B., Goldstein, H. & Gilbert, R. (2012). Opening the black box of record linkage. *Journal of Epidemiology and Community Health (1979),* 66**,** 1198-1198.

Hatton, D. C., Harrison-Hohner, J., Coste, S., Dorato, V., Curet, L. B. & Mccarron, D. A. (2005). Symptoms of postpartum depression and breastfeeding. *Journal of Human Lactation,* 21**,** 444-9; quiz 450-4.

Haynes, A. S., Derrick, G. E., Chapman, S., Redman, S., Hall, W. D., Gillespie, J. and Sturk, H. (2011) From 'our world' to the 'real world': exploring the views and behaviour of policy-influential Australian public health researchers. *Social Science & Medicine* 72(7): 1047–1055.

Health Research Authority. (2019). *Applying a proportionate approach to the process of seeking consent: HRA guidance* [*online*]. Available at: https://s3.eu-west-2.amazonaws.com/www.hra.nhs.uk/media/documents/Proportionate_approach_to_seeking_consent_HRA_Guidance.pdf [Accessed 14th September 2023].

Health Research Authority and Medicine and Healthcare Regulatory Products Agency (2018). *Joint statement on seeking consent by electronic methods*. [Online]. Available at: HRA and MHRA publish joint statement on seeking and documenting consent using electronic methods (eConsent) - Health Research Authority [Accessed 14th September 2023].

Health and Safety Executive. (2023). *Council services and activities* [Online]. Available at: https://www.hse.gov.uk/services/localgovernment/services-activities.htm [Accessed 8th August 2023].

Health Data Research UK. (2023). *Health data research explained* [Online]. Available at: https://www.hdruk.org/about-us/what-we-do/health-data-research-explained/ [Accessed 7th August 2023].

Health Research Authority. (2023). *Public Involvement* [Online]. Available at: https://www.hra.nhs.uk/planning-and-improving-research/best-practice/public-involvement/ [Accessed 9th February 2023].

Heclo, H. (1974) *Modern Social Politics in Britain and Sweden*. New Haven: Yale University Press.

Henderson, H., Bridges, S., Bryant, M., Pickett, K., Wilsdon, J. (2022a). *Using linked data as a local health intelligence tool for early life health: A qualitative study protocol. figshare*. Dataset. https://doi.org/10.6084/m9.figshare.20331618.v1

Henderson, H., Bridges, S., Padgett, L., Pickett, K., Wilsdon, J. (2022b). *The use of linked data evidence in UK decision-making related to early life health: Systematic mapping protocol*. figshare. Dataset. https://doi.org/10.6084/m9.figshare.20109593.v1

Henderson, J., Carson, C., Redshaw, M. (2016). Impact of preterm birth on maternal well-being and women's perceptions of their baby: a population-based survey. *British Medical Journal Open,* 6(10)**,** e012676. [Online]. Available at: https://bmjopen.bmj.com/content/6/10/e012676 [Accessed 4th Jan 2019].

Hindes, I. (2022) Born and Bred in (BaBi) East London Workshop: Summary report. Unpublished.

Hirst, C. (2022) Born and Bred in (BaBi) Wakefield Priority-Setting Workshop 16th November 2021: A Summary of Findings. Unpublished.

Hodgkinson, S., Beers, L., Southammakosane, C. & Lewin, A. (2014). Addressing the mental health needs of pregnant and parenting adolescents. *Pediatrics,* 133**,** 114-22.

Hodgkinson, S. C., Colantuoni, E., Roberts, D., Berg-Cross, L. & Belcher, H. M. (2010). Depressive symptoms and birth outcomes among pregnant teenagers. *Journal of Pediatric Adolescent Gynecology,* 23**,** 16-22.

Hopf, Y. M., Bond, C., Francis, J., Haughney, J. & Helms, P. J. (2014). 'The more you link, the more you risk ...' - A focus group study exploring views about data linkage for pharmacovigilance. *British Journal of Clinical Pharmacology,* 78**,** 1143-1150.

Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X. (2013). *Applied logistic regression, third edition,* Hoboken, NJ.

House Of Commons and Health and Social Care Committee. (2019). *First 1000 days of life: Thirteenth Report of Session 2017-19* [Online]. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmhealth/1496/1496.pdf [Accessed 15th February 2022].

Hughes, R. (1998) Considering the Vignette Technique and its Application to a Study of Drug Injecting and HIV Risk and Safer behaviour, *Sociology of Health and Illness* 20 (3) pp.381- 400.

Ibanez, G., Bernard, J. Y., Rondet, C., Peyre, H., Forhan, A., Kaminski, M., Saurel-Cubizolles, M.-J. & Group, E. M.-C. C. S. (2015). Effects of Antenatal Maternal Depression and Anxiety on Children's Early Cognitive Development: A Prospective Cohort Study. *PloS One,* 10**,** e0135849.

Immergut, E. M. (1998) The theoretical core of the new institutionalism. *Politics and Society* 26(1):5–34.

Imperial College London. (2021). *Urgent vs Important Matrix* [Online]. Available at: https://www.imperial.ac.uk/media/imperial-college/administration-and-support-services/staff-development/public/impex/Prioritisation-and-Time-Managment.pdf][Accessed 24 March 2021].

Innvær, S., Vist, G., Trommald, M. and Oxman, A. (2002) Health policy-makers' perceptions of their use of evidence: a systematic review. *Journal of Health Services Research & Policy* 7(4):239–244.

Insan, N., Weke, A., Rankin, J. & Forrest, S. (2022). Perceptions and attitudes around perinatal mental health in Bangladesh, India and Pakistan: a systematic review of qualitative data. *BioMed Central Pregnancy and Childbirth,* 22**,** 293.

Institute of Health Visiting. (2021). *0-19 Research Network (CRN Yorkshire & Humber).* [Online]. Available at: https://ihv.org.uk/news-and-views/news/0-19-research-network-crn-yorkshire-humber/ [Accessed 26 March 2021].

INVOLVE. (2012) Briefing notes for researchers: involving the public in NHS, public health and social care research. INVOLVE, Eastleigh. [Accessed 20 December 2020].

INVOLVE. (2018) *National Standards for Public Involvement*. [Online]. INVOLVE. Available at: https://www.invo.org.uk/wp-content/uploads/2019/02/71110_A4_Public_Involvement_Standards_v4_WEB.pdf [Accessed 24 March 2021].

Irani, E. (2019). The Use of Videoconferencing for Qualitative Interviewing: Opportunities, Challenges, and Considerations. *Clinical Nursing Research*, *28*(1), 3-8. https://doi.org/10.1177/1054773818803170

Isaacs, G. (2020). *A map of resources for co-producing research in health and social care.* [Online]. NIHR. Available at: https://learningforinvolvement.org.uk/wp-content/uploads/gravity_forms/6-341881053f9095fe395a1f986cd7085c/2020/06/Map-of-resources-Web-version-v1.2.pdf [Accessed 24 March 2021]

Jacob, C. M., Cooper, C., Baird, J. & Hanson, M. (2019). What quantitative and qualitative methods have been developed to measure the implementation of a life-course approach in public health policies at the national level? Copenhagen: WHO Regional Office for Europe.

James Bell Associates. (2018). *Guide to Data-Driven Decision Making: Using Data to Inform Practice and Policy Decisions in Child Welfare Organizations.* Washington, DC: Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.

James Lind Alliance Priority Setting Partnerships. (2022). *Final priority setting method* [Online]. Available at: https://www.jla.nihr.ac.uk/jla-guidebook/chapter-8/final-priority-setting-method.htm [Accessed 8th June 2023].

James, K. L., Randall, N. P. & Haddaway, N. R. (2016). A methodology for systematic mapping in environmental sciences. *Environmental Evidence,* 5**,** 7.

Jones, J. & Hunter, D. 1995. Consensus Methods for Medical And Health Services Research. *British Medical Journal,* 311, 376-380.

Kalkman, S., Van Delden, J., Banerjee, A., Tyl, B., Mostert, M. & Van Thiel, G. (2022). Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *Journal of Medical Ethics,* 48**,** 3-13.

Kapci, E. G., Kucuker, S. and Uslu, R. I. (2010). How Applicable Are Ages and Stages Questionnaires for Use With Turkish Children? *Topics in early childhood special education*, 30 (3), pp.176–188. [Online]. Available at: doi:10.1177/0271121410373149.

Katz, A., et al. (2021). Closing the loop: From system-based data to evidence-influenced policy and practice. *International Journal of Population Data Science,* 6(3) 1701.

Keith, J., Grimm, F. & Steventon, A. (2022). *How better use of data can help address key challenges facing the NHS* [Online]. The Health Foundation. Available at: https://www.health.org.uk/publications/long-reads/how-better-use-of-data-can-help-address-key-challenges-facing-the-nhs [Accessed 28th May 2023].

Kemal, Ö. (2020). Power Analysis and Sample Size, When and Why? *Turkish Archives of Otorhinolaryngol* ogy58(1) 3-4.

Kendall, S., Nash, A., Braun, A., Bastug, G., Rougeaux, E., & Bedford. H. (2014). Evaluating the use of a population measure of child development in the Healthy Child Programme Two Year Review.: UCL. [Online]. Available at: https://discovery.ucl.ac.uk/id/eprint/1493007/1/Kendall%20et%20al%20Evaluating%20child%20development%202014.pdf [Accessed 12 August 2022].

Khan, L. (2015). *Falling through the gaps: perinatal mental health and general practice* [Online]. Available at: https://maternalmentalhealthalliance.org/wp-content/uploads/RCGP-Report-Falling-through-the-gaps-PMH-and-general-practice-March-2015.pdf [Accessed 12th August 2022].

Kiernan, K. E. & Huerta, M. C. (2008). Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood. *British Journal of Sociology,* 59**,** 783-806.

Kim, C., *et al,.* (2018). A research utilisation framework for informing global health and development policies and programmes. *Health Research Policy & Systems* 16(1) 9.

Kingston, D., Tough, S. & Whitfield, H. (2012). Prenatal and Postpartum Maternal Psychological Distress and Infant Development: A Systematic Review. *Child Psychiatry & Human Development,* 43**,** 683-714.

Kinney, D. K. et al. (2008). Autism Prevalence Following Prenatal Exposure to Hurricanes and Tropical Storms in Louisiana. *Journal of Autism and Developmental Disorders*, 38 (3), pp.481–488. [Online]. Available at: doi:10.1007/s10803-007-0414-0.[Accessed 9th February 2024].

Kitson, A. L., *et al.,* (2008). Evaluating the successful implementation of evidence into practice using the PARiHS framework: theoretical and practical challenges. *Implementation Sci*ence,3 1.

Krol, K. M. & Grossmann, T. (2018). Psychological effects of breastfeeding on children and mothers. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz,* 61**,** 977-985.

Lancaster, C. A., Gold, K. J., Flynn, H. A., Yoo, H., Marcus, S. M. & Davis, M. M. (2010). Risk factors for depressive symptoms during pregnancy: a systematic review. *American Journal of Obstetrics & Gynecology,* 202**,** 5-14.

Langham, J. et al. (2023). Obstetric and neonatal outcomes in pregnant women with and without a history of specialist mental health care: a national population-based cohort study using linked routinely collected data in England. *The Lancet Psychiatry*, 10 (10), pp.748–759. [Online]. Available at: doi:10.1016/s2215-0366(23)00200-6.

Langhoff-Roos, J., Krebs, L., Klungsøyr, K., Bjarnadottir, R. I., Källén, K., Tapper, A.-M., Jakobsson, M., Børdahl, P. E., Lindqvist, P. G., Gottvall, K., Colmorn, L. B. & Gissler, M. (2014). The Nordic medical birth registers – a potential goldmine for clinical research. *Acta Obstetrician et Gynecologica Scandinavica,* 93**,** 132-137.

Latour, B. (2005) *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.

Lavis, J. N. (2006) Research, public policymaking, and knowledge-translation processes: Canadian efforts to build bridges. *The Journal of Continuing Education in the Health Professions,* 26(1): 37–45.

Lavis, J., Davies, H., Oxman, A., Denis, J. L., Golden-Biddle, K. & Ferlie, E. (2005). Towards systematic reviews that inform health care management and policy-making. *Journal of Health Services Research & Policy,* 10 Suppl 1**,** 35-48.

Lewis, J., & McNaughton Nicholls, C. (2014). Design Issues. In Richie, J., Lewis, J., McNaughton Nicholls. C., & Ormston, R. (Eds) *Qualitative research practice: a guide for social science students and researchers* (Second edn). SAGE Publications Ltd

Lawlor, D. A., Andersen, A.-M. N. & Batty, G. D. (2009). Birth cohort studies: past, present and future. *International Journal of Epidemiology,* 38**,** 897-902.

Leis, J. A., Heron, J., Stuart, E. A. & Mendelson, T. (2012). Associations Between Depressive and Anxious Symptoms and Prenatal Alcohol Use. *Maternal and Child Health Journal,* 16**,** 1304-1311.

Lewsey, J. D., Leyland, A. H., Murray, G. D. & Boddy, F. A. (2000). Using routine data to complement and enhance the results of randomised controlled trials. *Health Technology Assessment,* 4**,** 1-55.

Lincoln, Y. S. and Guba, G. (2009). The Only Generalization Is: There Is No Generalization. In Gomm, R., Hammersley, M. & Foster, P. (eds.) *Case Study Method.* London: Sage.

Lindblom, C. (1959) The science of 'Muddling Through'. *Public Administration Review,* 19: 79–88.

Lindquist, E. A. (1990) The third community, policy inquiry, and social scientists, in S. Brooks and A.-G. Gagnon (eds) *Social Scientists, Policy, and the State*. New York: Praeger, 21–51.

Liverpool City Council. (2021a). *Families Programme National Impact Study* [Online]. Available at: https://liverpool.gov.uk/media/9191/families-programme-national-impact-study-dclg.pdf [Accessed 12th June 2021].

Liverpool City Council. (2021b). *Liverpool Families Programme* [Online]. Available at: https://ehd.liverpool.gov.uk/kb5/liverpool/fsd/service.page?id=g46L-yV6SSg [Accessed 15th June 2021].

Lobe, B., Morgan, D., & Hoffman, K. A. (2020). Qualitative Data Collection in an Era of Social Distancing. *International Journal of Qualitative Methods*, *19*, 160940692093787. https://doi.org/10.1177/1609406920937875.

Local Government Association. (2023). *How are decisions made?* [Online]. Available at: https://www.local.gov.uk/our-support/councillor-and-officer-development/councillor-hub/introduction-local-government/how-are [Accessed 17th August 2023].

Lomas, J. (2000) Using 'linkage and exchange' to move research into policy at a Canadian foundation. *Health Affairs,* 19(3): 236–240.

Low, J. (2019). A Pragmatic Definition of the Concept of Theoretical Saturation. *Sociological Focus,* 52**,** 131-139.

Lugg-Widger, F. V., Angel, L., Cannings-John, R., Hood, K., Hughes, K., Moody, G. & Robling, M. (2018). Challenges in accessing routinely collected data from multiple providers in the UK for primary studies: Managing the morass. *International Journal of Population Data Science,* 3**,** 432.

Lyons, R. A., Jones, K. H., John, G., Brooks, C. J., Verplancke, J.-P., Ford, D. V., Brown, G. & Leake, K. (2009). The SAIL databank: linking multiple health and social care datasets. *BioMed Central Medical Informatics and Decision Making,* 9**,** 3-3.

Masefield, S., Willan, K., Darwin, Z., Blower, S., Nekitsing, C. & Dickerson, J. (2022) Bad data is worse than no data: limitations in publicly available routine perinatal mental health data for estimating prevalence and area variation associated with health inequalities. Unpublished.

Macfarlane, A., Dattani, N., Gibson, R., Harper, G., Martin, P., Scanlon, M., Newburn, M. & Cortina-Borja, M. (2019). Births and their outcomes by time, day and year: a

retrospective birth cohort data linkage study. *Health Services and Delivery Research,* 7**,** 1-268.

Mackinnon, D. P. & Luecken, L. J. (2008). How and for whom? Mediation and moderation in health psychology. *Health Psychology,* 27**,** S99-s100.

Malterud, K., Siersma, V. D. and Guassora, A. D. (2016) 'Sample Size in Qualitative Interview Studies: Guided by Information Power', *Qualitative Health Research*, 26(13), pp. 1753–1760. doi: 10.1177/1049732315617444.

MatCHNet. (2021). MatCHNet [Online]. Available at: https://matchnet.sphsu.gla.ac.uk/research/ [Accessed 8th April 2021].

Marmot, M. *et al.,* (2010). The Marmot review: Fair society, healthy lives. *The Strategic Review of Health Inequalities in England Post-2010*.

Marryat, L., Auyeung, B. & Speyer, L. (2022). Neonatal and child health and developmental outcomes in Scotland., *International Journal of Population Data Science*, 7(3). doi: 10.23889/ijpds.v7i3.2010.

Marsh, D. and Rhodes, R. A. W. (1992) *Policy Networks in British Government*. Oxford: Oxford University Press.

Marsh, D. and Smith, M. (2000) Understanding policy networks: towards a dialectical approach. *Political studies* 48(1): 4–21.

Marvin-Dowell, K. (2021) Born and Bred in Doncaster: Outputs from the stakeholder engagement workshop 10th June 2021. Unpublished.

Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews. Forum: *Qualitative Social Research,* 11, Article 8.

Maxwell, J. A. (2012). *A realist approach for qualitative research.* Thousand Oaks, SAGE Publications.

Mayne, J. (2008). Contribution Analysis: An Approach to Exploring Cause and Effect.*ILAC Brief*. ILAC Institutional learning and Change Institute.

McGowan, J.E., Alderdice, F.A., Holmes, V.A., Johnston, L. (2011). Early childhood development of late- preterm infants: a systematic review. *Pediatrics,* 127(6)**,** 1111-24.

Mc Grath-Lone, L. Jay, M. A., Blackburn, R., Gordon, E., Zylbersztejn, A., Wiljaars, L. & Gilbert, R. (2022). What makes administrative data "research-ready"? A systematic review and thematic analysis of published literature. *International Journal of Population Data Science,* 7**,** 1718.

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia Medica (Zagreb),* 22**,** 276-82.

McLennan, D. (2018). Data quality issues in administrative data. *In* E. Mackey and M. Elliot eds., ADRN.

Meagher, L., Lyall, C. & Nutley, S. (2008). Flows of knowledge, expertise and influence: A method for assessing policy and practice impacts from social science research. *Research Evaluation,* 17(3)**,** 163-173. doi:10.3152/095820208X331720

Meshberg-Cohen, S. & Svikis, D. (2007). Panic disorder, trait anxiety, and alcohol use in pregnant and nonpregnant women. *Comprehensive Psychiatry,* 48**,** 504-10.

Mikkelsen, B., Williams, J., Rakovac, I., Wickramasinghe, K., Hennis, A., Shin, H.-R., Farmer, M., Weber, M., Berdzuli, N., Borges, C., Huber, M. & Breda, J. (2019). Life course approach to prevention and control of non-communicable diseases. *British Medical Journal Clinical research ed.),* 364**,** l257-l257.

Mind. (2022). *Postnatal depression and perinatal mental health* [Online]. Available at: https://www.mind.org.uk/information-support/types-of-mental-health-problems/postnatal-depression-and-perinatal-mental-health/about-maternal-mental-health-problems/ [Accessed 25th May 2022].

Ministry of Housing Communities and Local Government (2021). *Local Data Accelerator Fund for Children and Families: Prospectus.* Ministry Of Housing Communities and Local Government. [Online]. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/973289/Prospectus_-_Local_Data_Accelerator_Fund.pdf [Accessed 10 June 2021].

Ministry of Housing Communities and Local Government. (2019a) *National Evaluation of the Troubled Families Programme 2015 – 2020: Family Outcomes – national and local datasets, Part 4* [Online]. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/786891/National_evaluation_of_the_Troubled_Families_Programme_2015_to_2020_family_outcomes___national_and_local_datasets_part_4.pdf?_ga=2.222715675.1230623744.1566890831-26383294.1545319507 [Accessed 15th February 2022].

Ministry of Housing Communities and Local Government (2019b). *The English Indices of Deprivation 2019.* Ministry Of Housing Communities and Local Government. [Online]. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/853811/IoD2019_FAQ_v4.pdf [Accessed 20 June 2023].

Mitton, C., Adair, C. E., McKenzie, E., Patten, S. B. and Waye Perry, B. (2007) Knowledge transfer and exchange: review and synthesis of the literature. *The Milbank Quarterly* 85(4): 729–768.

Mizrahi, E. M. (2014). Neonatal Seizures. *In* M. J. Aminoff and R. B. Daroff eds. *Encyclopedia of the Neurological Sciences (Second Edition)*. Oxford, Academic Press. 316-321.

Moher, D et al., (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Physical Therapy*. **89**(9) pp.873-880.

Molas-Gallart, J et al. (2000) Assessing the Non-academic Impact of Grant-funded Socio-economic Research: Results from a Pilot Study, *Research Evaluation,* 9/3:171-82.

Moreno-Giménez, A., Campos-Berga, L., Nowak, A., Sahuquillo-Leal, R., D'ocon, A., Hervás, D., Navalón, P., Vento, M. & García-Blanco, A. (2021). Impact of maternal age on infants' emotional regulation and psychomotor development. *Psychological Medicine*, 1-12.

Morton, S. (2015). Progressing research impact assessment: A 'contributions' approach. *Research Evaluation,* 24(4) 405-419.

Mughal, M. K., Giallo, R., Arnold, P. D., Kehler, H., Bright, K., Benzies, K., Wajid, A. & Kingston, D. (2019). Trajectories of maternal distress and risk of child developmental delays: Findings from the All Our Families (AOF) pregnancy cohort. *Journal of Affective Disorders,* 248**,** 1-12.

Muraca, G. M. & Joseph, K. S. (2014). The Association Between Maternal Age and Depression. *Journal of Obstetrics and Gynaecology Canada,* 36**,** 803-810.

Nasreen, H.-E. *et al.,* (2013). Impact of maternal depressive symptoms and infant temperament on early infant growth and motor development: Results from a population based study in Bangladesh. *Journal of affective disorders*, 146 (2), pp.254-261. [Online]. Available at: doi:10.1016/j.jad.2012.09.013. [Accessed 9th February 2024].

National Institute for Health and Care Research. (2023). *Call for proposals: NIHR Health Determinants Research Collaborations (HDRC) specification document* [Online]. Available at: https://www.nihr.ac.uk/documents/specification-document-call-for-proposals-nihr-health-determinants-research-collaborations-hdrc/32383 [Accessed 4th April 2023].

National Institute for Health and Care Excellence. (2014) Antenatal and postnatal mental health: clinical management and service guidance [Online]*.* Available at: *https://www.nice.org.uk/guidance/cg192* (Accessed: 26th May 2022).

Newcastle Hospitals NHS Foundation Trust. (2023). *Neonatal intensive care unit* [Online]. Available at: https://www.newcastle-hospitals.nhs.uk/services/maternity/after-your-baby-is-born/neonatal-intensive-care-unit/ [Accessed 20th June 2023].

Newsom, J. T. (2021). *Sample Size Issues for Categorical Analyses and Logistic Regression* [Online]. Available at: https://web.pdx.edu/~newsomj/cdaclass/ho_sample%20size.pdf [Accessed 1st August 2023].

NHS. (2023a). *Maximising the benefits of research: Guidance for integrated care systems* [Online]. Available at: https://www.england.nhs.uk/long-read/maximising-the-benefits-of-research/ [Accessed 4th April 2023].

NHS. (2023b). *Video conferencing with colleagues* [Online]. Available at: https://transform.england.nhs.uk/information-governance/guidance/using-video-conferencing-tools-communicate-colleagues/ [Accessed 6th April 2023].

NHS England. (2023). NHS Data Model and Dictionary: DELIVERY METHOD CODE [Online]. Available at: https://www.datadictionary.nhs.uk/data_elements/delivery_method_code.html?hl=delivery%2Cmethod [Accessed 23rd August 2023].

NHS.UK. (2022). *Perinatal mental health* [Online]. Available at: https://www.england.nhs.uk/mental-health/perinatal/ [Accessed 25th May 2022].

NHS. (2019). *The NHS Long Term Plan* [Online]. Available at: https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/ [Accessed 29th May 2023].

NHS Digital. (2022). *Read Codes* [Online]. Available at: https://digital.nhs.uk/services/terminology-and-classifications/read-codes [Accessed 7th August 2023].

NHS England (2023). About the Maternity Services Data Set. [Online]. Available at: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-

sets/maternity-services-data-set/about-the-maternity-services-data-set [Accessed 19th February 2024].

NIHR. (2023). *Call for proposals: NIHR Health Determinants Research Collaborations (HDRC) specification document* [Online]. Available at: https://www.nihr.ac.uk/documents/specification-document-call-for-proposals-nihr-health-determinants-research-collaborations-hdrc/32383 [Accessed 30th May 2023].

NIHR. (2022). *£50 million awarded to local government to tackle interventions for health inequalities through research* [Online]. Available at: https://www.nihr.ac.uk/news/50-million-awarded-to-local-government-to-tackle-interventions-for-health-inequalities-through-research/31654 [Accessed 30th May 2023].

NIHR (2020a). *NIHR launches new Centre for Engagement and dissemination* [Online]. Available at: https://www.nihr.ac.uk/news/nihr-launches-new-centre-for-engagement-and-dissemination/24576 [Accessed 24 March 2021].

NIHR (2020b). *The James Lind alliance Guidebook: version 9*. [Online]. Available at: https://www.jla.nihr.ac.uk/jla-guidebook/downloads/JLA-Guidebook-V9-download-March-2020.pdf [Accessed 24 March 2021].

NIHR Applied Research Collaboration South London. (2021). *Approved digital linkages* [Online]. Available at: https://www.arc-sl.nihr.ac.uk/research-and-implementation/our-research-methods/applied-informatics/applied-informatics-hub-3 [Accessed 20th May 2021].

NIHR Guy's and St Thomas' BRC. (2023). *The eLIXIR Programme* [Online]. Available at: http://www.guysandstthomasbrc.nihr.ac.uk/microsites/elixir/about-the-programme/ [Accessed 8th August 2023].

Northern Ireland Longitudinal Study Research Support Unit. (2021). *About the NILS-RSU Project* [Online]. Available at: https://www.nils-rsu.co.uk/about-us/ [Accessed 12th June 2021].

Nuffield Trust. (2022) *Linking datasets for the benefit of children and young people.* [Online]. Available at: https://www.nuffieldtrust.org.uk/event/data-linkage-for-the-benefit-of-children-outcomes-and-services [Accessed 28th February 2022].

Nuffield Trust. (2021). *Low birth weight* [Online]. Available at: https://www.nuffieldtrust.org.uk/resource/low-birth-weight [Accessed 3rd August 2022].

Nutley, S., Walter, I. C. and Davies, H. T. O. (2007) *Using Evidence: How Research Can Inform Public Services.* Bristol: Policy Press.

Nutley, S., Davies, H. & Walter, I. (2003). Evidence-based policy and practice: cross-sector lessons from the United Kingdom. *Social Policy Journal of New Zealand*, 29-48.

O'Connor, T. G., Heron, J., Golding, J., Beveridge, M. & Glover, V. (2002). Maternal antenatal anxiety and children's behavioural/emotional problems at 4 years. Report from the Avon Longitudinal Study of Parents and Children. *British Journal of Psychiatry,* 180, 502-8.

O'Connor, M. J. & Whaley, S. E. (2006). Health care provider advice and risk factors associated with alcohol consumption following pregnancy recognition. *Journal of Studies on Alcohol and Drugs,* 67, 22-31.

OECD (2019), *Health in the 21st Century: Putting Data to Work for Stronger Health Systems*, OECD Health Policy Studies, OECD Publishing, Paris, https://doi.org/10.1787/e3b23f8e-en.

Office for Health Improvement and Disparities. (2023). *Breastfeeding at 6 to 8 weeks after birth: quarterly data for 2022 to 2023* [Online]. Available at: https://www.gov.uk/government/statistics/breastfeeding-at-6-to-8-weeks-after-birth-quarterly-data-for-2022-to-2023 [Accessed 1st August 2023].

Office for National Statistics. (2021). *ONS Longitudinal Study* [Online]. Available at: https://www.ons.gov.uk/aboutus/whatwedo/paidservices/longitudinalstudyls [Accessed 12th June 2021].

Oldroyd, J. (2005). Low birth weight in South Asian babies in Britain: time to reduce the inequalities*. Fukushima Journal of Medical Science*, 51 1-10.

Oliver, K. & Boaz, A. (2019). Transforming evidence for policy and practice: creating space for new conversations. *Palgrave Communications,* 5**,** 60.

Oliver, K. & Cairney, P. (2019). The dos and don'ts of influencing policy: a systematic review of advice to academics. *Palgrave Communications,* 5**,** 21.

Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BioMed Central Health Services Research*, 14. https://doi.org/10.1186/1472-6963-14-2

Oliver, K., Kothari, A., & Mays, N. (2019). The dark side of coproduction: Do the costs outweigh the benefits for health research? *Health Research Policy and Systems*, *17* (1), 1-10. https://doi.org/10.1186/s12961-019-0432-3

Olsen, J., Melbye, M., Olsen, S. F., Sørensen, T. I., Aaby, P., Andersen, A. M., Taxbøl, D., Hansen, K. D., Juhl, M., Schow, T. B., Sørensen, H. T., Andresen, J., Mortensen, E. L., Olesen, A. W. & Søndergaard, C. (2001). The Danish National Birth Cohort--its background, structure and aim. *Scandinavian Journal of Public Health,* 29**,** 300-7.

Orton, L., Lloyd-Williams, F., Taylor-Robinson, D., O'Flaherty, M., & Capewell, S. (2011). The use of research evidence in public health decision making processes: Systematic review. *PloS One,* 6(7), E21704.

Oxford University Hospitals NHS Foundation Trust. (2023). *Research and Development: About us* [Online]. Available: https://www.ouh.nhs.uk/researchers/about/ [Accessed 29th March 2023].

Padmanabhan, S., Carty, L., Cameron, E., Ghosh, R. E., Williams, R. & Strongman, H. (2019). Approach to record linkage of primary care data from Clinical Practice Research Datalink to other health-related patient data: overview and implications. *European Journal of Epidemiology,* 34**,** 91-99.

Page, M. J, McKenzie. J. E, Bossuyt, P. M, Boutron, I, Hoffmann, T.C, Mulrow, C.D, *et al*. (2020) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *British Medical Journal, 2021*;372:n71. doi: 10.1136/bmj.n71.

Papadopoulos, A., Nichols, E. S., Mohsenzadeh, Y., Giroux, I., Mottola, M. F., Van Lieshout, R. J., & Duerden, E. G. (2022). Prenatal and postpartum maternal mental health and neonatal motor outcomes during the COVID-19 pandemic. *Journal of affective disorders reports*, 10, 100387. https://doi.org/10.1016/j.jadr.2022.100387. [Accessed 9th February 2024].

Parkhurst. (2017). *The Politics of Evidence (Open Access): From evidence-based policy to good governance of evidence.* Routledge. doi: 10.4324/9781315675008

Pawson, R. (2006). *Evidence-based Policy: A Realist Perspective.* London: SAGE.

Petticrew, M., & Roberts, Helen. (2006). *Systematic reviews in the social sciences: a practical guide / Mark Petticrew and Helen Roberts.* Malden, MA; Oxford: Blackwell.

Pettinger, K. J., *et al.* (2020). Starting school: educational development as a function of age of entry and prematurity. *Archives of Disease in Childhood,* 105(2) 160.

Pilkington, R., Motegomerie, A., Grant, J., Gialamas, A., Malvaso, C., Smithers, L., Chittleborough, C., & Lynch, J. (2019). *An innovative linked data platform to improve the wellbeing of children– the South Australian Early Childhood Data Project* [[Online]. Available at: https://www.aihw.gov.au/getmedia/75c9f692-c72c-438d-b68b-91318f780154/Australias-Welfare-Chapter-8-summary-18Sept2019.pdf.aspx [Accessed 7th August 2023].

Pocock, T., Smith, M. & Wiles, J. (2021). Recommendations for Virtual Qualitative Health Research During a Pandemic. *Qualitative Health Research,* 31**,** 2403-2413.

Polidano, C., Zhu, A. & Bornstein, J. C. (2017). The relation between cesarean birth and child cognitive development. *Scientific Reports,* 7**,** 11483.

Powell, T., Gheera, M., Foster, D., Long, R. & Kennedy, S. (2021). *Early intervention: policy and provision*, UK Parliament Research Briefings.

Prady, S. L., Darwin, Z., Nekitsing, C., Blower, S., Willan, K. & Dickerson, J. (2022) Inequalities in identification and management of perinatal mental health problems: A review of academic and local reports. Unpublished.

Prady, S. L., Endacott, C., Dickerson, J., Bywater, T. J. & Blower, S. L. (2021). Inequalities in the identification and management of common mental disorders in the perinatal period: An equity focused re-analysis of a systematic review. *PloS One,* 16**,** e0248631.

Prady, S. L., Pickett, K. E., Petherick, E. S., Gilbody, S., Croudace, T., Mason, D., Sheldon, T. A. & Wright, J. (2016). Evaluation of ethnic disparities in detection of depression and anxiety in primary care during the maternal period: Combined analysis of routine and cohort data. *British Journal of Psychiatry,* 208**,** 453-461.

Prady S. L., Pickett K. E., Petherick E. S., Gilbody S., Croudace T., Mason D., Sheldon, T. A, & Wright, J. (2015) Presentation to primary care with depression and anxiety during the maternal period: a combined analysis of routine and cohort data. *British Journal of Psychiatry* (in press)

Public Health England. (2021a). *Public health outcomes framework: Child and Maternal Health* [Online]. Available at: https://fingertips.phe.org.uk/profile/child-health-profiles/data#page/1/gid/1938133223/pat/15/ati/6/are/E12000003/iid/20401/age/173/sex/2/cid/4/tbm/1 [Accessed 24 March 2021].

Public Health England (2021b). *Early years high impact area 2: Supporting maternal and family mental health. In:* GOV.UK.

Pybus, K, *et al.,* (2023). Falling through the gaps: A mixed methods study to understand the barriers to recording and capture of perinatal mental health data. Unpublished.

Qresearch. (2021). *Qresearch: GENERATING NEW KNOWLEDGE TO IMPROVE PATIENT CARE* [Online]. Available at: https://www.qresearch.org [Accessed 17th June 2021].

Quigley, M. A., Hockley, C., Carson, C., Kelly, Y., Renfrew, M. J. & Sacker, A. (2012). Breastfeeding is Associated with Improved Child Cognitive Development: A Population-Based Cohort Study. *The Journal of Pediatrics,* 160**,** 25-32.

Raftery, J., Roderick, P. & Stevens, A. (2005). Potential use of routine databases in health technology assessment. *Health Technology Assessment,* 9**,** 1-92, iii-iv.

Raynor, P., *et al.* (2008). Born in Bradford, a cohort study of babies born in Bradford, and their parents: Protocol for the recruitment phase. *BioMed Central Public Health*, 8, 1-13. https://doi.org/10.1186/1471-2458-8-327

Regenstrief Institute. (2023). *Indiana Network for Population Health focuses on social determinants of health* [*online*]. Available at: https://www.regenstrief.org/article/indiana-network-for-population-health-social-determinants-of-health/#:~:text=The%20Indiana%20Network%20for%20Population,information%20on%20social%20determinants%20of [Accessed 8th September 2023].

Regulation (EU) 2016/679 (General Data Protection Regulation). *Official Journal of the European Union*, L119, 1-88.

Rein, M. (1980) Methodology for the study of the interplay between social science and social policy. *International Social Science Journal,* xxii(2): 361-368.

Relton, C., et al. (2010). Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. *British Medical Journal,* 340 c1066

Richardson, C., Akhtar, I., Smith, C., Edmondson, A., Morris, A., Hargreaves, J., Rhodes, C., & Taylor, J. (2019). Effective involvement: A report on the evaluation of a research awareness training package for public involvement in health research. *Research Involvement and Engagement*, *5* (21), 1-10. https://doi.org/10.1186/s40900-019-0151-5

Riches, N., Coleman, A., Gadsby, E. & Peckham, S. (2015). The Role of Local Authorities in health issues: A policy document analysis. PRUComm.

Ritchie, J., Lewis, J., McNaughton Nicholls, C. and Ormston, R., (2014). *Qualitative research practice: a guide for social science students and researchers*. 2nd ed. London: SAGE Publications Ltd.

Robling, M., Lugg-Widger, F., Cannings-John, R., Sanders, J., Angel, L., Channon, S., Fitzsimmons, D., Hood, K., Kenkre, J., Moody, G., Owen-Jones, E., Pockett, R., Segrott, J. & Slater, T. (2021). The Family Nurse Partnership to reduce maltreatment and improve child health and development in young children: the BB:2-6 routine data-linkage follow-up to earlier RCT. *Public Health Research,* 9**,** 1-160.

Rudan, I., Gibson, J. L., Ameratunga, S., El Arifeen, S., Bhutta, Z. A., Black, M., Black, R. E., Brown, K. H., Campbell, H., Carneiro, I., Chan, K. Y., Chandramohan, D., Chopra, M., Cousens, S., Darmstadt, G. L., Gardner, J. M., Hess, S. Y., Hyder, A. A., ... Webster, J. (2008). Setting Priorities in Global Child Health Research Investments: Guidelines for Implementation of CHNRI Method. *Croatian Medical Journal*, *49*(6), 720-733. https://doi.org/10.3325/cmj.2008.49.720

Russell, T. (1997). *Pair wise ranking made easy* [Online]. PLA Notes, IIED London. Available at: https://www.iied.org/sites/default/files/pdfs/migrate/G01675.pdf [Accessed 12th June 2023].

Sacchi, C. *et al.,* (2018). Does infant negative emotionality moderate the effect of maternal depression on motor development? *Early Human Development*, 119, pp.56–61. [Online]. Available at: doi:10.1016/j.earlhumdev.2018.03.006. [Accessed 9th February 2024].

Santos, I. S., Matijasevich, A., Domingues, M. R., Barros, A. J. & Barros, F. C. (2010). Long-lasting maternal depression and child growth at 4 years of age: a cohort study. *J Journal of Pediatrics,* 157**,** 401-6.

Saunders, L. (2005) Research and policy: reflections on their relationship. *Evidence & Policy* 1(3): 383–390.

Schildberg-Hoerisch, H. (2011). Does parental employment affect children's educational attainment? *Economics of Education Review*, Elsevier, vol. 30(6),1456-1467.

Schober, P. & Vetter, T. R. (2021). Logistic Regression in Medical Research. *Anesthesia and analgesia,* 132**,** 365-366.

Schofield, J. W. (2009). Increasing the Generalizability of Qualitative Research. In Gomm, R., Hammersley, M. and Foster, P. (eds.) *Case Study Method.* London: Sage.

Scottish Government. (2012a). *A Scotland-wide Data Linkage Framework for Statistics and Research: Consultation Paper on the Aims and Guiding Principles: Consultation Analysis* [Online]. Available at: https://www.gov.scot/binaries/content/documents/govscot/publications/consultation-analysis/2012/08/scotland-wide-data-linkage-framework-statistics-research-consultation-analysis/documents/00400972-pdf/00400972-pdf/govscot%3Adocument/00400972.pdf [Accessed 9th February 2023].

Scottish Government (2012b). *Joined up data for better decisions: a strategy for improving data access and analysis*. The Scottish Government. [Online]. Available at: https://www.gov.scot/binaries/content/documents/govscot/publications/strategy-plan/2012/11/joined-up-data-better-decisions-strategy-improving-data-access-analysis/documents/joined-up-data-better-decisions-strategy-improving-data-access-analysis/joined-up-data-better-decisions-strategy-improving-data-access-analysis/govscot%3Adocument/00408151.pdf [Accessed 15th February 2022].

ScottisH Informatics Programme. (2011). *A Blueprint for Health Records Research in Scotland: Appendix 1.* [Online]. Available at: https://www.scot-ship.ac.uk/sites/default/files/Reports/Appendix_1.pdf [Accessed 11th June 2021].

Scottish Longitudinal Study Development & Support Unit. (2021). *Introduction to the Scottish Longitudinal Study* [Online]. Available at: https://sls.lscs.ac.uk/about/ [Accessed 12th June 2021].

Selvin, S. (1996). *Statistical analysis of epidemiologic data,* 2nd ed. New York Oxford, Oxford University Press.

Shapiro, J. R., Klein, S. L. & Morgan, R. (2021). Stop 'controlling' for sex and gender in global health research. *British Medical Journal Global Health,* 6**,** e005714.

SILC. (2021). *How It Works* [Online]. Available at: http://www.datalinkagescotland.co.uk/how-it-works [Accessed 12th June 2021].

Simcock, G., Kildea, S., Elgbeili, G., Laplante, D. P., Stapleton, H., Cobham, V. & King, S. (2016). Age-related changes in the effects of stress in pregnancy on infant motor development by maternal report: The Queensland Flood Study. *Development Psychobiology,* 58**,** 640-59.

Singer, L. T., Salvator, A., Guo, S., Collin, M., Lilien, L. & Baley, J. (1999). Maternal Psychological Distress and Parenting Stress After the Birth of a Very Low-Birth-Weight Infant. *JAMA,* 281**,** 799-805.

Slomian, J., Honvo, G., Emonts, P., Reginster, J. Y. & Bruyère, O. (2019). Consequences of maternal postpartum depression: A systematic review of maternal and infant outcomes. *Women's Health, (London),* 15**,** 1745506519844044.

Smith, K. E. (2013). Beyond Evidence Based Policy in Public Health: The Interplay of Ideas. *Palgrave Communications.* https://doi.org/10.1057/9781137026583

Smith, K. E. (2007). Health inequalities in Scotland and England: The Contrasting Journeys of Ideas from Research into Policy, *Social Science and Medicine*, 64/7:1438-39.

Sohal K, Mason D, Birkinshaw J *et al.,* (2022) Connected Bradford: A Whole System Data Linkage Accelerator [version 2; peer review: 2 approved]. *Wellcome Open Res*, 7:26 (https://doi.org/10.12688/wellcomeopenres.17526.1)

Spencer, M. E. (1973) What is charisma? *The British Journal of Sociology* 24(3): 341-354.

Squires, J., et al. (2009). ASQ-3 Technical Report [Online]: Brookes Publishing. Available at: https://agesandstages.com/wp-content/uploads/2017/05/ASQ-3-Technical-Appendix_web.pdf [Accessed 23rd August 2023].

Stake, R. E. (2009). The Case Study Method in Social Inquiry. In Gomm, R., Hammersley, M. & Foster, P. (eds.) *Case Study Method.* London: Sage.

Stake, R. E. (1995). *The Art of Case Study Research*: SAGE Publications.

Stake, R. E. (1978). The Case Study Method in Social Inquiry. *Educational Researcher* 7(2) 5-8.

Staley, K. (2009). Exploring Impact: public involvement in NHS, public health and social care research**.** INVOLVE, Eastleigh.

Stansfield, C., Dickson, K. & Bangpan, M. (2016). Exploring issues in the conduct of website searching and other online sources for systematic reviews: how can we be systematic? *Systematic Reviews,* 5**,** 191.

Stephens, S., Ford, E., Paudyal, P. & Smith, H. (2016). Effectiveness of Psychological Interventions for Postnatal Depression in Primary Care: A Meta-Analysis. *Annals Family Medicine,* 14**,** 463-72.

Suh, I., Nam, C. M. & Kang, H. G. (1996). A Study on Misclassification Arising from Random Error in Exposure Measurement. *Epidemiology & Health,* 18**,** 108-118.

Sutcliffe, A. G., Barnes, J., Belsky, J., Gardiner, J. & Melhuish, E. (2012). The health and development of children born to older mothers in the United Kingdom: observational study using longitudinal cohort data. *British Medical Journal,* 345**,** e5116.

Syddall, H. E., Simmonds, S. J., Carter, S. A., Robinson, S. M., Dennison, E. M. & Cooper, C. (2019). The Hertfordshire Cohort Study: an overview [version 1 peer review: 3 approved]. *F1000 research,* 8**,** 82-82.

Taylor, R. (2023). *Capturing the impact of nursing research: VICTOR and the role of the RCN* [Online]. Available at: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjzgeGyyu2AAxXjVkEAHSbKB5AQFnoECB8QAQ&url=https%3A%2F%2Fwww.rcn.org.uk%2F-%2Fmedia%2FRoyal-College-Of-Nursing%2FDocuments%2FProfessional-Development%2FResearch%2F2019-research-conference%2FFringe-3.pdf&usg=AOvVaw3IR2Feq9DKU-LHTULxOihY&opi=89978449. [Accessed 21st August 2023].

Tenny, S. & Hoffman, M, R. (2023) Odds Ratio. [Online]. Available at: https://www.ncbi.nlm.nih.gov/books/NBK431098/#:~:text=The%20confidence%20interval%20gives%20an,ratio%20found%20in%20the%20sample. [Accessed 23rd August 2023].

The King's Fund. (2022a). *How does the NHS in England work and how is it changing?* [Online]. Available at: https://www.kingsfund.org.uk/audio-video/how-does-nhs-in-england-work [Accessed 30th May 2023].

The King's Fund. (2022b). *Integrated care systems: how will they work under the Health and Care Act?* [Online]. Available at: https://www.kingsfund.org.uk/audio-video/integrated-care-systems-health-and-care-act [Accessed 30th May 2023].

The National Lottery Community Fund. (2022). *Data-informed decision making: Insights from The National Lottery Community Fund's A Better Start Programme* [Online]. Available at: https://www.tnlcommunityfund.org.uk/media/insights/documents/Programme-Insight-5-data-informed-decision-making-FINAL.pdf [Accessed 29th May 2023].

The University of Edinburgh. (2021a). *Generation Scotland* [Online]. Available at: https://www.ed.ac.uk/generation-scotland/who-we-are/about [Accessed 12th June 2021].

The University of Edinburgh. (2021b). *Scottish Health and Ethnicity Linkage Study* [Online]. Available at: https://www.ed.ac.uk/usher/scottish-health-ethnicity-linkage [Accessed 12th June 2021].

Tingay, K. S., et al. (2019). Record linkage to enhance consented cohort and routinely collected health data from a UK birth cohort. *International Journal of Population Data Science* 4(1).

Todd, O. M., Burton, J. K., Dodds, R. M., Hollinghurst, J., Lyons, R. A., Quinn, T. J., Schneider, A., Walesby, K. E., Wilkinson, C., Conroy, S., Gale, C. P., Hall, M., Walters, K., & Clegg, A. P. (2020). New Horizons in the use of routine data for ageing research. *Age and ageing*, 49(5), 716–722. https://doi.org/10.1093/ageing/afaa018.

Tong, A., Synnot, A., Crowe, S., Hill, S., Matus, A., Scholes-Robertson, N., Oliver, S., Cowan, K., Nasser, M., Bhaumik, S., Gutman, T., Baumgart, A., & Craig, J. C. (2019). Reporting guideline for priority setting of health research (REPRISE). *BioMed Central Medical Research Methodology*, *19*(1), 1–11. https://doi.org/10.1186/s12874-019-0889-3.

Tweed E, Cimova K, Craig P*, et al.* (2022). P71 Unlocking data to inform public health policy and practice: decision-maker perspectives on the use of cross-sectoral data as

part of a whole-systems approach. *Journal of Epidemiology and Community Health,* 76, A78.

UKRI. (2023). *Guidance on engaging the public with your research* [Online]. Available at: https://www.ukri.org/manage-your-award/good-research-resource-hub/guidance-on-engaging-the-public-with-your-research/ [Accessed 13th September 2023].

UKRI. (2022a). *How Research England supports research excellence* [Online]. Available at: https://www.ukri.org/about-us/research-england/research-excellence/ref-impact/ [Accessed 29th May 2023].

UKRI. (2022b). *Longitudinal Administrative Data Spine Scoping Project Grant For The SPF UK Population Lab Wave I* [Online]. Available at: https://gtr.ukri.org/projects?ref=ES%2FS016732%2F1 [Accessed 22nd November 2022].

UKRI. (2021). *Wirral Child Health and Development Study* [Online]. Available at: https://mrc.ukri.org/research/facilities-and-resources-for-researchers/cohort-directory/wirral-child-health-and-development-study/ [Accessed 15th June 2021].

Understanding Patient Data. (2018). *Why an opt-out rather than an opt-in or consent?* [Online]. Available at: https://understandingpatientdata.org.uk/news/why-an-opt-out [Accessed 29th May 2023].

University of Birmingham. (2023). *New £1.4 million partnership to advance maternal and infant health research using health data* [Online]. Available at: https://www.birmingham.ac.uk/news/2023/new-1.4-million-partnership-to-advance-maternal-and-infant-health-research-using-health-data [Accessed 29th May 2023].

University of Bristol (2021a). *Avon Longitudinal Study of Parents and Children.* [Online]. University of Bristol Available at: http://www.bristol.ac.uk/alspac/researchers/publications/2019-a-m/. [Accessed 11th June 2021].

University of Bristol. (2021b). *Linkage to routine health and social records* [Online]. Available at: http://www.bristol.ac.uk/alspac/researchers/our-data/linkage/ [Accessed 12th June 2021].

University College London. (2021). *Research and Policy Briefs* [Online]. Available at: https://www.ucl.ac.uk/epidemiology-health-care/research/epidemiology-and-public-health/research/health-and-social-surveys-research-group/studies-55 [Accessed 17th June 2021].

University of Exeter. (2022). *Grey Literature: What is Grey Literature?* [Online]. Available at: https://libguides.exeter.ac.uk/c.php?g=670055&p=4756572 [Accessed 15th February 2022].

University of Liverpool. (2023). *Children Growing up in Liverpool (C-GULL)* [Online]. Available at: https://www.liverpool.ac.uk/children-growing-up-in-liverpool/research/ [Accessed 8th August 2023].

University of Liverpool. (2021). *Wirral Child Health and Development Study: First Steps* [Online]. Available: at: https://www.liverpool.ac.uk/population-health/research/groups/first-steps/about/ [Accessed 15th June 2021].

University of York. (2023). *Video conferencing: Zoom* [Online]. Available at: https://www.york.ac.uk/it-services/services/zoom/#tab-4 [Accessed 11th August 2023].

Upadhyay, R. P., Naik, G., Choudhary, T. S., Chowdhury, R., Taneja, S., Bhandari, N., Martines, J. C., Bahl, R. & Bhan, M. K. (2019). Cognitive and motor outcomes in children born low birth weight: a systematic review and meta-analysis of studies from South Asia. *BioMed Central Pediatrics,* 19**,** 35.

Vezyridis, P. and Timmons, S. (2017). Understanding the care.data conundrum: New information flows for economic growth, *Big Data & Society*. doi: 10.1177/2053951716688490.

Voit, F. A. C., Kajantie, E., Lemola, S., Räikkönen, K., Wolke, D., & Schnitzlein, D. D. (2022). Maternal mental health and adverse birth outcomes. *PloS one,* 17(8), e0272210. https://doi.org/10.1371/journal.pone.0272210

Waind, E. (2020). Trust, security and public interest: striking the balance A narrative review of previous literature on public attitudes towards the sharing, linking and use of administrative data for research. *International Journal of Population Data Science,* 5**,** 1368.

Waind, E. & Mc Grath-Lone, L. (2021). *The Education & Child Health Insights from Linked Data (ECHILD) Database Key messages from stakeholders*, ADR UK.

Walder, D. J. et al. (2014). Prenatal maternal stress predicts autism traits in 6½ year-old children: Project Ice Storm. *Psychiatry research,* 219 (2), pp.353–360. [Online]. Available at: doi:10.1016/j.psychres.2014.04.034. [Accessed 9th February 2024].

Waldfogel, J., Han, W. J. & Brooks-Gunn, J. (2002). The effects of early maternal employment on child cognitive development. *Demography,* 39**,** 369-92.

Walt, G. (1994) How far does research influence policy? *European Journal of Public Health* 4(4): 233–235.

Walter, I., Nutley, S. and Davis, H. (2005) What works to promote evidence-based practice? *Evidence & Policy* 1(3): 335–363.

Walters, K., Rait, G., Griffin, M., Buszewicz, M. & Nazareth, I. (2012). Recent trends in the incidence of anxiety diagnoses and symptoms in primary care. *PloS One,* 7**,** e41670-e41670.

Ward, V., A. House and S. Hamer (2009). Developing a framework for transferring knowledge into action: a thematic analysis of the literature. *Journal of Health Services Research & Policy* 14(3) 156-164.

Warren-Gash, C. (2017). *Linking and sharing routine health data for research in England* [Online]. PHG Foundation. Available at: https://www.phgfoundation.org/media/119/download/PHGF-linking-and-sharing-routine-health-data-for-research-in-england.pdf?v=1&inline=1 [Accessed 28th May 2023].

Watson, H. et al. (2019). A systematic review of ethnic minority women's experiences of perinatal mental health conditions and services in Europe. *PloS One*, 14 (1), pp.e0210587–e0210587. [Online]. Available at: doi:10.1371/journal.pone.0210587.

Weiss, C. (1990) The uneasy partnership endures: social science and govern- ment, in S. Brooks and A.-G. Gagnon (eds) *Social Scientists, Policy, and the State*. New York: Praeger: 97–111.

Weiss, C. (1982) Policy research in the context of diffuse decision making. *The Journal of Higher Education* 53(6): 619–639.

Weiss, C. (1979) The many meanings of research utilization. *Public Administration Review* 39(5): 426-431.

Weiss, C. (1977) Research for policy's sake: the enlightenment function of social research. *Policy Analysis,* 3: 531-547.

Williams, C. J., Turner, K. M., Burns, A., Evans, J. & Bennert, K. (2016). Midwives and women's views on using UK recommended depression case finding questions in antenatal care. *Midwifery,* 35**,** 39-46.

Wellcome. (2023). *Using an engaged research approach* [Online]. Available at: https://wellcome.org/grant-funding/guidance/using-engaged-research-approach#main [Accessed 13th September 2023].

Wellcome Trust. (2015). *Enabling Data Linkage to Maximise the Value of Public Health Research Data: full report* [Online]. Available at: https://cms.wellcome.org/sites/default/files/enabling-data-linkage-to-maximise-value-of-public-health-research-data-phrdf-mar15.pdf [Accessed 2023].

Welsh Government. (2019) *Analysis of Flying Start outcomes using linked data: childcare analysis - Initial research findings produced by analysing childcare data from the Flying Start programme linked with other health and education data.* [Online]. Available at: https://gov.wales/sites/default/files/pdf-versions/2021/1/3/1610564495/analysis-flying-start-outcomes-using-linked-data-childcare-analysis.pdf [Accessed 15th February 2022].

West, J., et al. (2013). UK-born Pakistani-origin infants are relatively more adipose than white British infants: findings from 8704 mother-offspring pairs in the Born-in-Bradford prospective birth cohort. *Journal of Epidemiology and Community Health* 67(7) 544.

West Yorkshire Health and Care Partnership. (2023). *Five Year Plan - Chapter 1: Introduction* [Online]. Available at: https://www.wypartnership.co.uk/publications/our-five-year-plan/five-year-plan-introduction [Accessed 30th May 2023].

Willan, K., *et al.*, (2022). *Report 3 Reducing Inequalities in Perinatal Mental Health Care: Exploration of inequalities in identification and treatment of perinatal mental health concerns: A description of key similarities and differences across the West Yorkshire region*. [Online]. Available at: https://eprints.whiterose.ac.uk/201017/1/Report_3_Data_Report_V1_April_22.pdf [Accessed 23rd August 2023).

Williams Lea. (2016). *National Data Guardian for Health and Care: Review of Data Security, Consent and Opt-Outs* [Online]. National Data Guardian. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF [Accessed 7th August 2023].

Wilsdon, J., Curry, S., and Gadd, E. (2023). *REF 2028: A clap for the Frap* [Online]. Research Professional News Available at: https://www.researchprofessionalnews.com/rr-news-political-science-blog-2023-6-a-clap-for-the-frap/ [Accessed 7th August 2023].

Wilson, K. M., Brady, T. J., and Lesesne. C (2011). An organizing framework for translation in public health: the Knowledge to Action Framework. *Preventing Chronic Disease,* 8(2) A46.

World Health Organisation. (2014). *Providing health intelligence to meet local needs: A practical guide to serving local and urban communities through public health observatories* [Online]. Available at:

https://apps.who.int/iris/bitstream/handle/10665/152645/9789241508162_eng.pdf?sequence=1&isAllowed=y [Accessed 5th June 2023].

World Health Organization. (2008). Closing the gap in a generation: health equity through action on the social determinants of health - Final report of the commission on social determinants of health.

Wright, J., McEachan, R., & Mathai, M. (2021). Why is the Born in Bradford cohort study important for child health? *Archives of Disease in Childhood*, 1-2. https://doi.org/10.1136/archdischild-2020-321231

Wright, J., Small, N., Raynor, P., Tuffnell, D., Bhopal, R., Cameron, N., Fairley, L., A Lawlor, D., Parslow, R., Petherick, E. S., Pickett, K. E., Waiblinger, D., & West, J. (2013). Cohort profile: The Born in Bradford multi-ethnic family cohort study. *International Journal of Epidemiology*, 42(4), 978-991. https://doi.org/10.1093/ije/dys112

Wright, J., Wood, M., Dezateux, C., Scammel, K., Harper, G., West, J., Muckle, S., Shore, R Sohal, K., Mason, D.,Mon-Williams, M.,Lockyer, B., Twite, S., Homer, K; Watmuff, A., Cronin, M.,Firman, N., Cooper, D., & Sheldon, T. (2022) Whole system data linkage accelerators: a North-South partnership to unlock public health data. Unpublished.

Yuen, M., Hall, O. J., Masters, G. A., Nephew, B. C., Carr, C., Leung, K., Griffen, A., Mcintyre, L., Byatt, N. & Moore Simas, T. A. (2022). The Effects of Breastfeeding on Maternal Mental Health: A Systematic Review. *Journal of Women's Health (Larchmt),* 31, 787-807.

Zachariadis, M., Scott, S. and Barrett, M. (2013). Methodological Implications of Critical Realism for Mixed-Methods Research. *Management Informations Systems Quarterly,* 37(3), 855-879.

Zavez, A. et al. (2021). Delivery Mode and Child Development at 20 Months of Age and 7 Years of Age in the Republic of Seychelles. *Maternal and Child Health Journal,* 25 (12), pp.1930–1938. [Online]. Available at: doi:10.1007/s10995-021-03239-8.