Estimating Lifetime Effects of Child Development for Economic Evaluation:

An Exploration of Methods and their Application to a Population Screen for Postnatal Depression

Richard David Mattock

PhD

University of York

Health Sciences

November 2018

Abstract

Background: Early health interventions affecting child development can subsequently influence lifetime health and economic outcomes. These lifetime effects may be excluded from economic evaluation as empirical evidence covering the required time horizon is rarely available. One example is screening for postnatal depression where current guidelines do not account for lifetime effects despite evidence of a detrimental association between maternal depression and child development.

Aims: To develop a methodological approach to estimate lifetime effects for economic evaluation and determine their influence on an evaluation assessing the cost-effectiveness of postnatal depression screening.

Methods: Lifetime effects are estimated by linking results from two empirical studies. Firstly, growth curve models establish the effects of postnatal depression on development measures for children aged 3-11 using data from the Millennium Cohort Study. Secondly, child development measures are entered as explanatory variables in linear regression models predicting effects on lifetime health and economic outcomes using data from the 1970 British Cohort Study. An economic evaluation is conducted for scenarios which exclude/include lifetime effects to determine their influence on cost-effectiveness results.

Findings: Postnatal depression was detrimentally associated with children's cognitive and socioemotional development up to age 11. Detrimental changes in cognitive and socioemotional development were negatively associated with lifetime outcomes. Postnatal depression exposure was predicted to reduce children's lifetime Quality Adjusted Life Years, increase healthcare and crime costs, and generate fewer monetary returns in education and employment. Cost-effectiveness results changed when including lifetime effects, leading to the recommendation of a screening strategy which treats a greater proportion of depressed mothers.

Conclusions: Lifetime effects can influence cost-effectiveness results and their exclusion risks providing a partial analysis. This research demonstrates methods to estimate and include lifetime effects in economic evaluation. Similar approaches could be applied elsewhere to provide additional evidence for economic evaluation of other childhood interventions.

List of Contents

List of Tables

Chapter 3:

Chapter 4:

Chapter 5:

Chapter 6:

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

Chapter 6:

Chapter 7:

Acknowledgements

I would like to express immense gratitude to my supervisors Professor Karen Bloor and Professor Gerry Richardson for their help and guidance throughout the PhD process. I would also like to acknowledge the members of my thesis advisory panel, Professor Kate Pickett and Professor Martin Chalkley, who provided useful and timely advice. Finally, I would like to thank my family and friends for the encouragement and support they have offered over the past years.

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.

Chapter 1: Introduction

1.1 Summary

Healthcare allocation decisions are commonly informed though economic evaluation, a type of analysis that ranks competing interventions according to their associated costs and benefits. This thesis addresses a challenge faced specifically in the economic evaluation of child health interventions: some interventions delivered early in life can affect child development and may have subsequent costs and benefits on health and social outcomes across the lifespan. Such lifetime effects are expected to occur throughout adulthood, spanning several decades after the administration of a childhood intervention. Consequently, lifetime effects are difficult to measure empirically and are often excluded from analyses. The purpose of this thesis is to investigate methods to include lifetime effects in economic evaluation, and further, to demonstrate their importance in decision making.

The overarching aims of this research are:

> To develop and describe a methodology for estimating lifetime effects.
>
> To determine the influence of lifetime effects on the cost-effectiveness results in an applied, UK based, economic evaluation.

This introductory chapter further details the rationale for the thesis.

1.2 Context: Health Decision Making and Economic Evaluation Literature

1.2.1 Health Interventions and Healthcare Allocation Decisions

Health interventions often require consumption of expensive resources which may include use of medical devices, medicines, vaccines, one-on-one time with highly trained specialists and accommodation in hospital facilities. In publicly funded health systems, such as the UK National Health Service (NHS), it is not possible to adopt every intervention achieving a health benefit as the cost of doing so would far exceed the total amount of money society would be prepared to pay (Eichler et al., 2004). Therefore, adoption decisions are required that determine an appropriate distribution of the finite budget for healthcare (Angelis et al., 2017), (Drummond et al., 2015). Each intervention adopted by a health system has an associated opportunity cost as healthcare resources are consumed which could have been spent to achieve health benefits

elsewhere (Palmer and Raftery, 1999). As such, the allocation of healthcare resources leads to changes in levels of morbidity and mortality for members of society.

1.2.2 Economic Evaluation to inform Health Decision Making

UK health decision making is commonly advised or informed by economic evaluation; an analytical framework that can provide answers to difficult questions of resource allocation. Fundamentally, economic evaluation is a comparative analysis that establishes the best course of action based on the costs and consequences associated with different health interventions (Drummond et al., 2015). To inform healthcare allocation decisions adequately, economic evaluations use socially desirable *decision rules* to rank competing health interventions according to their costs and consequences (Drummond et al., 2015).

A key feature of economic evaluation is the use of evidence to support decision making (Sculpher et al., 2006). Evidence is required that identifies the costs and consequences associated with competing health interventions and this is measured on *decision endpoints* – the outcomes of interest for decision makers. The most reliable sources of evidence are usually obtained empirically through primary analysis of data obtained from clinical trials or observational study designs (Cooper et al., 2007). In the absence of empirical data, useful information might be obtained from other less reliable sources which may include expert opinion (Soares et al., 2018).

The exact type of evidence required varies for each analysis depending on the specifics of the decision problem and the adopted analytical framework. *Trial based evaluations* are single clinical studies that split patients into different treatment groups with each receiving a competing health intervention or a relevant comparator e.g. treatment as usual. Evidence is obtained in a trial-based framework by observing the decision endpoints achieved for each group (Hawkins et al., 2012). Alternatively, economic evaluations can be conducted using *decision analytic models* (DAMs) which represent the decision problem and disease progression as different health/disease and exposure states. Evidence is required to inform many factors (or parameters) in a DAM, including estimates of the decision endpoints associated with different health/disease/exposure states within the model (Cooper et al., 2007).

1.2.3 Lifetime Horizon in the Economic Evaluation of Child Health Interventions

Regardless of the framework selected for economic evaluation, evidence should be obtained over a time horizon which can account appropriately for all the important costs and benefits of the health interventions under investigation (Sculpher et al., 2006). In many cases, a *lifetime*

horizon is most appropriate as the influence of health interventions can often continue across the lifespan and influence outcomes such as mortality (Drummond et al., 2015).

A lifetime horizon is likely to be required when establishing the costs and consequences associated with a subset of health interventions delivered during childhood as these interventions have the potential to influence a child's later lifespan development.[1] The process of lifespan development is the extraordinary change that occurs within all individuals as they progress from cellular organisms to increasingly complex infants, children, teenagers and finally mature adults, determining how individuals look, think, behave and interact with the world around them (Santrock, 2003). Health interventions affecting the development process could impact substantially on the individual throughout adulthood and might influence factors relevant for health and social decision makers e.g. lifestyle choices, family circumstances, education, employment, health behaviours, criminal activity (Halfon et al., 2014) (Heckman, 2006), (Ben-Shlomo and Kuh, 2002).

1.2.4 Direct Sources of Evidence for Economic Evaluation

The adoption of a lifetime horizon poses an analytical problem when conducting economic evaluation of early childhood health interventions as long-term evidence is required across a time horizon spanning several decades. The usual approach to obtaining empirical evidence is by *directly estimating* the relationship by observing the outcomes achieved in different treatment/disease/exposure groups, as is illustrated in Figure 1.1, (Hawkins et al., 2012).

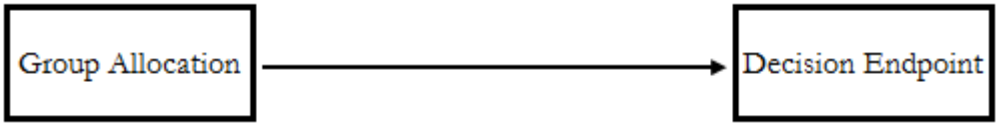


Figure 1.1: Depicts the method of direct estimation. Note that group allocation might include treatment, exposure, or disease groups versus a relevant control.

Randomised controlled trials (RCTs) are often considered the best, or "gold standard", method when identifying direct evidence as they eliminate several potential biases through the random allocation of participants to relevant groups (Akobeng, 2005). For pragmatic reasons, RCTs are

[1] Theoretical and empirical evidence is reviewed in chapter two which explains the link between childhood health interventions and lifespan development.

unlikely to be appropriate when directly estimating lifetime evidence for the economic evaluation of childhood health interventions. RCTs are associated with high costs (Bothwell et al., 2016) and are therefore unlikely to be commissioned if studies require a very lengthy time horizon. Even if commissioned, RCT evidence would not become available until far into the future when children had sufficiently progressed across their lifespan. Additionally, RCTs might be considered unethical if the required evidence relates to the effects of disease/health/exposure states rather than comparisons between different treatment groups i.e. children ought not to be *knowingly* assigned harmful stimuli (Relton et al., 2010).

If RCTs are inappropriate, direct evidence might be obtained from observational studies which can be more ethically viable because harmful stimuli are experienced through natural phenomena rather than being purposely assigned in trials (Webb et al., 2016). Longitudinal birth cohorts are likely to provide the most relevant observational study design when identifying the lifetime effects associated with child health interventions – this type of study design obtains information on a group of new-borns and follows them as they progress towards adulthood by collecting information in repeated data sweeps.

Several birth cohort studies have been commissioned in the UK and continue to follow up participants into later adulthood. These include the 1946 National Survey of Health and Development (Wadsworth et al., 2005), the 1958 National Child Development Study (Power and Elliott, 2006) and the 1970 British Cohort Study (Elliott and Shepherd, 2006). Whilst it is possible that long term evidence could be obtained from longitudinal birth cohorts, researchers are reliant on a relatively small number of historical studies which may not have collected the variables required for the specific objectives of research.

1.2.5 Indirect Sources of Evidence for Economic Evaluation

Alternatively, relevant evidence could be obtained through *indirect estimation,* where group effects are observed on some intermediate outcome measure(s) that are then translated to the final endpoints of interest. The key assumption made when conducting indirect estimation is that the intermediate outcomes necessarily occur along the pathway of effect between treatment/disease/exposure and the final endpoint of interest, (Figure 1.2.). A common type of indirect estimation is the use of surrogate outcomes as substitutes for final outcomes in clinical trials (Fleming and Powers, 2012). For instance, the effects of cardiovascular drugs on reducing final endpoints such as stroke or myocardial infarction are sometimes measured in terms of their effects in reducing the surrogate endpoint blood pressure (Aronson, 2005).

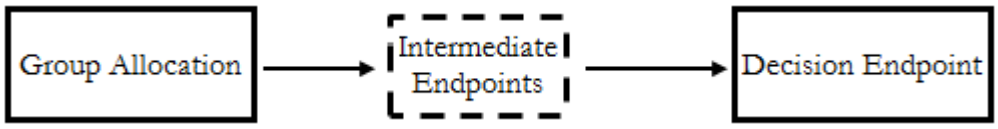


Figure 1.2: Depicts the method of indirect estimation.

Indirect estimation requires empirical evidence to be linked from at least two studies. As with direct estimation the most appropriate sources of evidence to identify observed group effects on intermediate endpoints are likely to be obtained from RCTs and/or observational study designs. A variety of different techniques might be appropriate when establishing how intermediate effects translate to effects on final decision endpoints. For example, direct evidence between group allocation and final decision endpoint might be used to quantify this relationship as is required when validating surrogate endpoints (Fleming and Powers, 2012); the pathway might be represented using a mathematical model whose parameters are informed through empirical data (Buxton et al., 1997); or it could be assumed that the relationship is appropriately represented by a mathematical function or distribution (Garnett et al., 2011).[2]

The benefit of indirect estimation is to increase the feasibility of obtaining evidence as it reduces the time horizon required for the empirical studies. A reduced time horizon could also mean more appropriate measures are available as these studies are less likely to rely on historical data collections. The use of indirect evidence in economic evaluation is demonstrated by the Washington State Institute for Public Policy (WSIPP) (2017) who make recommendations regarding the cost-benefit of several early childhood social policies, and by Hummel et al. (2011) when assessing the cost-effectiveness of early years public health programmes. In both examples, indirect evidence facilitated the conduct of economic evaluation where no direct empirical evidence was available.

Whilst the evaluations by WSIPP (2017) and Hummel et al. (2011) demonstrate the potential to inform economic evaluation through indirectly estimated evidence, the primary purpose of these studies was to inform policy and neither study provides a detailed justification regarding the appropriateness of the methods adopted. In addition, the studies are not reported in enough detail to enable other analysts to replicate their methodologies.

[2] Chapter four provides a detailed discussion regarding the different techniques that might be applied when translating observed intermediate effects to final lifetime decision endpoints.

1.2.6 Gap in the Existing Evidence

There is a sparsity of published literature transparently discussing appropriate methods for estimating the lifetime effects of childhood health interventions for economic evaluation. Limitations in data availability can leave decision makers uninformed (Soares et al., 2013), and may impact on policy recommendations if the size of lifetime effects is substantial enough to affect the results of economic evaluation. Indirect estimation might provide the most feasible route to obtaining lifetime evidence and future research developing these methods could facilitate decisions in the allocation of child health interventions.

1.3 Research Aims

This thesis addresses methods in the economic evaluation of childhood health interventions where analyses may be limited due to the unavailability of evidence across the lifetime.

The aims of this research are to:

1. Develop and describe a methodology which indirectly estimates the lifetime effects of early childhood circumstances by linking results across two empirical analyses.
2. Determine whether lifetime estimates influence the cost-effectiveness results in an applied, UK based, economic evaluation.
3. Discuss the appropriateness of indirect estimation when estimating lifetime decision endpoints for economic evaluation.

This research could be used to inform the design of other studies and may provide additional and important evidence in the economic evaluation of child health interventions.

1.4 Research Methodology and Applied Example

The research objectives are addressed throughout this thesis by demonstrating methodologies within an applied economic evaluation as this provides a realistic test of how indirect estimation could work in practice. The applied economic evaluation assesses the cost-effectiveness of UK population screening strategies for postnatal depression.

Screening is a clinical method used to identify a sub group of a population who are at high risk of having a disease/disorder. Following a positive screen this sub group will usually proceed to further diagnostic tests and/or be offered an appropriate treatment (Gilbert et al., 2001). There is evidence to suggest that child development is detrimentally affected following exposure to

symptoms of postnatal depression (Kingston and Tough, 2014), (Sanger et al., 2015). However, published economic evaluations do not consider children's lifetime outcomes when assessing the cost-effectiveness of screening of mothers which may be due to a lack of direct evidence.

This research advances a method for indirectly estimating the effects of postnatal depression on children's lifetime decision endpoints and uses these estimates to conduct an economic evaluation assessing the cost-effectiveness of screening for postnatal depression. The evaluation updates and extends an existing NICE (2018) analysis by accounting for the costs and consequences of screening on children's lifetime decision endpoints. The influence of the lifetime outcomes on final policy recommendations is demonstrated by comparing the results of the applied economic evaluation with and without children's lifetime effects. Recommendations from the applied research could be used to update existing clinical guidelines regarding the appropriate screening strategy for postnatal depression in the UK.

1.5 Thesis Structure

Chapter two begins with a literature review summarising some of the background theory in the fields of lifespan development and health economics. The review explains why early intervention can affect development and speculates on the nature of these effects across the lifespan. A further objective of the chapter is to establish an appropriate philosophical framework for UK economic evaluation of childhood health interventions given their potential to affect lifespan development. Accordingly, the different philosophical approaches that can be taken when conducting economic evaluation are discussed. As the philosophical approach determines the type of evaluation, chapter two informs the design of the applied analyses subsequently used in the research by specifying the decision endpoints to be adopted – these being Quality Adjusted Life Years (QALYs), healthcare costs, monetised costs/returns incurred in sectors outside of health, and economic productivity.

The thesis continues with three empirical chapters estimating indirect lifetime evidence. Chapter three introduces the postnatal depression applied example and specifies the additional lifetime evidence on child development required when assessing the cost-effectiveness of screening. The purpose of the chapter is to demonstrate the first stage in the methodological approach when obtaining indirect lifetime evidence. A primary empirical analysis is conducted identifying the association between childhood exposure to symptoms of postnatal depression and cognitive and socioemotional development outcomes. The analysis uses time series data from the Millennium Cohort Study allowing intermediate effects to be identified at multiple time points for children aged between three and eleven years.

Chapter four focuses on identifying an appropriate method for extrapolating the observed intermediate effects to effects on lifetime decision endpoints. It is suggested that this second stage of indirect estimation could be achieved through mathematical modelling. A scoping review is conducted with the objective of identifying a mathematical model suitable for predicting lifetime decision endpoints for the applied example. Several models are identified which estimate lifetime effects from measures of child development, but none use outcomes that are fully relevant decision endpoints (i.e. Quality Adjusted Life Years /healthcare costs). To inform the further primary analyses in this thesis, a synthesis is performed across the identified studies establishing some common characteristics that should be adopted when using models to estimate the lifetime effects of child development.

In the absence of relevant evidence in the secondary literature, chapter five reports the results of a primary analysis which applies a mathematical model in data from the 1970 British Cohort Study. This analysis uses several parallel linear regression models to predict lifetime decision endpoints from measures of cognitive and socioemotional development when children were aged ten. The chapter combines results from the mathematical model with observed evidence from chapter three to estimate the incremental Quality Adjusted Life Years, healthcare costs and monetised returns to the crime and education sectors for children exposed to symptoms of postnatal depression, when compared with children who were not exposed. The specific role of the empirical chapters in the indirect estimation process is illustrated in Figure 1.3.

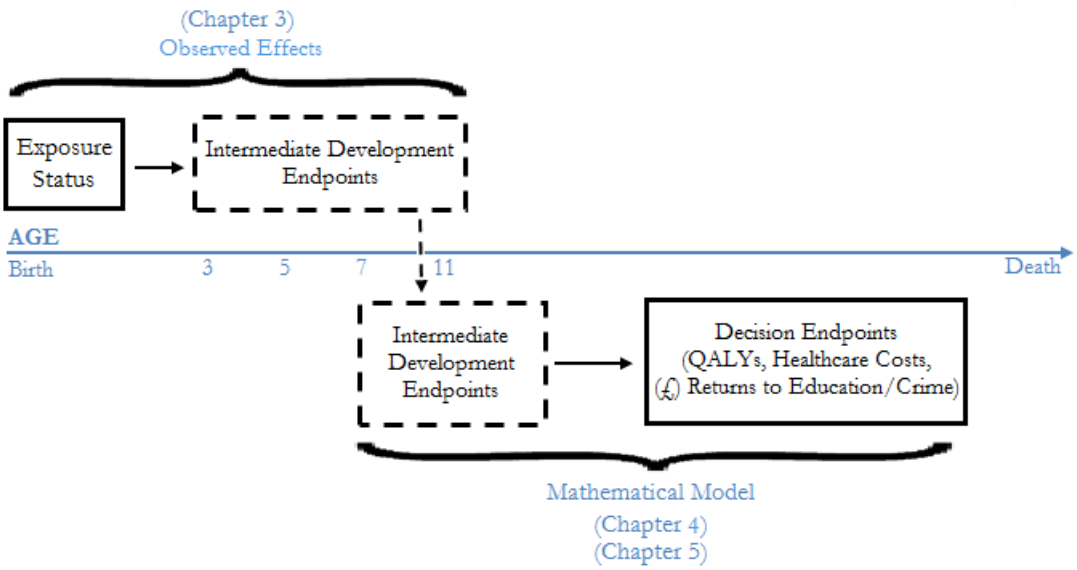


Figure 1.3: Depiction of the methodological approach for indirect estimation used in the applied example to estimate the lifetime effects associated with children exposed to symptoms of postnatal depression.

Chapter six describes the applied economic evaluation assessing the cost-effectiveness of screening for postnatal depression in the UK. This analysis is conducted with and without the previously estimated lifetime effects. The policy recommendations from the applied economic evaluation are discussed in full here. Additionally, a value of information analysis is conducted to establish future research priorities and to determine whether the acquisition of future lifetime evidence is likely to be cost-effective.

Finally, chapter seven discusses the methodological approach used in this research to indirectly estimate lifetime effects. The validity of and methodological extensions to indirect estimation are considered, along with alternative approaches that might be applied when estimating lifetime effects. The value of including lifetime effects and the cost-effectiveness of obtaining this information in the applied example is also discussed. Several recommendations are made that could aid other researchers when estimating lifetime effects in the economic evaluation of childhood health interventions elsewhere.

1.6 Boundaries of Research

This thesis does not consider several other research challenges potentially faced by analysts conducting economic evaluation of early childhood health interventions. Ungar (2009) provides a comprehensive summary of methodological challenges in child health economic evaluation including: the difficulty in measuring children's health outcomes, the extension of an intervention's effects beyond children and into family members, the patterns of health and disease in childhood, and the equity of childhood intervention.

The boundaries of the thesis are also restricted by considering only the appropriate methodologies for the economic evaluation of childhood *health* interventions. It is recognised that other social policies are likely to have effects on child development. The methodological approaches for *health* economic evaluation are not necessarily transferable when conducting an economic evaluation of a policy that is commissioned for by, say, the education or crime sectors.

Finally, the applied example in this research adapts a published economic evaluation (NICE, 2018) assessing the cost-effectiveness of screening for postnatal depression by including estimates regarding the lifetime effects associated with postnatal depression. There are several other research limitations associated with the published analysis. Whilst some of these limitations were accounted for it was not possible to address them *all* as the research was focused on achieving the primary aim of this thesis, which was to discuss the appropriate methodologies for estimating lifetime effects.

1.7 Key Terms

Health intervention and *health technology* are used interchangeably in this thesis. Both terms follow the definition provided by the World Health Organisation (WHO, 2018) which describes a health technology as "an application of organised knowledge and skills [that is] developed to solve a health problem or improve quality of lives".

The term "c*hildhood circumstance"* is used as an overarching concept that refers to childhood health technologies, childhood disease, childhood exposures, and any other event that may occur during childhood that has the potential to affect a child's later development.

Throughout the thesis the term *"lifetime effects"* is used to refer to the adulthood health and economic consequences of childhood circumstances which specifically occur through their effects to the lifespan development process.

Chapter 2: Methods for the Economic Evaluation of Childhood Health Technologies

2.1 Summary

This chapter presents a literature review summarising some of the background theory in the fields of lifespan development and UK health economic evaluation of childhood health technologies. The review explains why early intervention health technologies can affect child development and speculates on the nature of these effects across the lifespan. A further focus of the chapter is to establish an appropriate philosophical framework for UK economic evaluation of childhood health technologies given their potential to affect lifespan development; accordingly, the different philosophical approaches that can be taken when conducting economic evaluation are discussed. As the philosophical approach determines the type of evaluation, the chapter informs the design of the applied analyses following later in this thesis.

2.2 Objectives

In support of the overarching aims of the research in this thesis the objective of this chapter is:

> To conduct a literature review to inform the design of (i) an approach for the indirect estimation of lifetime effects; and (ii) a philosophical framework for economic evaluation given the potential lifetime effects associated with child health technologies.

The findings from the review are used to inform the two primary analyses in chapters three and five which indirectly estimate the lifetime effects for children exposed to symptoms of postnatal depression, and an economic evaluation assessing the cost-effectiveness of screening for postnatal depression in chapter six.

This chapter begins by reviewing some of the contemporary lifespan development literature describing what development is, the mechanisms that drive change, and the importance of early life. The review demonstrates that child health technologies have the potential to influence lifespan development if they target children's early environmental conditions. Theoretical and empirical evidence further suggests that, if they do influence the lifespan development process, early childhood health technologies may go on to affect a variety of other adulthood outcomes relevant to social policy makers. A method for indirect estimation is suggested that could account for the pathway between child health technologies and adulthood outcomes by establishing the effects of a health technology on intermediate outcomes measuring lifespan development.

The next section of the review describes the contrasting philosophical frameworks that can be adopted when conducting economic evaluation. The discussion considers which philosophical framework might be appropriate given the potential lifetime effects of child health technologies. The chapter also considers the appropriate analytical methods consistent with the suggested philosophical framework and the requirements of decision makers in the UK. Specifically, the chapter answers the following questions regarding methods used for economic evaluation of child health technologies:

1. What *type* of economic evaluation should be conducted?
2. What *decision rules* should be implemented?
3. Which *decision endpoints* should be used to measure costs and benefits?

2.3 Lifespan Development

2.3.1 Overview of Lifespan Development

The human lifespan follows a universal sequence beginning at conception, moving on to birth, maturation and reproduction before finally ending at death (Bogin and Smith, 1996). As individuals progress through this sequence they undergo unmistakable changes in physique, function, ability, emotion, behaviour, personality and in their general interaction with their surrounding environment (Santrock, 2003). The term lifespan development is the all-encompassing concept that describe all the intra-individual (i.e. within person) changes that occur over time (Molenaar et al., 2003), (Smith et al., 2015).

The extent of development is extraordinary when considering changes in structure and function that occur as individuals mature from simple cellular organisms to complex adult humans. Because of the vast subject area, development is sometimes studied independently as three comprehensible sub-domains (Berk, 2013): the physical domain accounts for the structural and organisational development of the individual, for example height, weight and neural circuitry; the cognitive domain refers to mental capacities such as information processing, language, memory and perception; while the socioemotional domain is concerned with emotion, personality and social interaction with other individuals (Keenan and Evans, 2009), (Berk, 2013).

Whilst studied independently, each development domain influences and is influenced by each sub-domain meaning the overarching process is interdependent (Gauvain and Parke, 2008). Infantile growth, for example, might lead to walking abilities (physical domain) which enable social interaction with others (socioemotional domain) that enhance later language faculties (cognitive domain) (Đorđić et al., 2016). Similarly, the growth and refinement of neural

networks result in the emergence of cognitive abilities such as logic and reasoning which may eventually influence behaviour (Battista et al., 2018).

The emergence of skills and capabilities from morphological changes describe a type of development known as transformational change (Overton, 2010b). A defining example of transformation change in nature occurs in the caterpillar-butterfly where the organism develops the capacity to fly through organisational changes in its structure i.e. wings (Overton, 2010b). Likewise, it is argued that human development may occur as a series of transformations or discontinuous changes (Boyd et al., 2014), (Newman and Newman, 2017), (Santrock, 2003). Advocates of discontinuous change point to the qualitative differences observed across prenatal, infant, child, adolescent and adulthood stages in the human lifespan (Lightfoot, 2013).

There is also strong evidence that development occurs continuously through small variational changes (Overton, 2010a). Continuous development might make sense to parents who see children accumulate skills through adaptive changes built upon over time. This is illustrated by the development of coordination and motor skills as a child slowly learns to crawl, stand, walk, run and jump (Santrock, 2003). While some reviewers emphasise the debate between continuous/discontinuous development (e.g. Boyd et al. (2014), Lightfoot (2013) & Santrock (2003)), it is the view of Overton (2010b) that development ought to and can only be fully considered in terms of *both* transformational (discontinuous) and variational (continuous) change.

2.3.2 The Mechanisms driving Human Development

The mechanisms producing development are a topic of much debate. Philosophers, including Plato and Descartes, have argued that a *natural* internal mechanism is responsible for producing change (Boyd et al., 2014). In support of the nature hypothesis, Mendel's research on inheritance discovered that internal biological molecules called *genes* explain characteristics shared between parents and offspring (Gottlieb, 1998). Genes have now been clarified as specific sequences of DNA that are contained in the cell nucleus and recent research has established the complete sequence of DNA that makes up the human genome (IHGSC, 2001).

Crick's (1958) theory, *the central dogma*, attributed the biological role of genes as the sole mechanism responsible for protein synthesis, a process that occurs through the joint procedures of transcription of DNA to RNA and translation of RNA to protein (Buxbaum, 2007). According to the central dogma, genes exist in a "vacuum" cut off from all intracellular and extracellular influences (Crick, 1958), (Gottlieb, 1998). The central dogma was an influential theory in molecular biology (Thieffry and Sarkar, 1998) where it became a popular assumption

that protein synthesis occurs in a unidirectional flow of information from gene to protein (Crick, 1958), (Gottlieb, 1998).

Given their essential biological properties (Buxbaum, 2007), and in accord with the central dogma, *predetermined epigenesis* supposed that proteins provide the basis for *structural maturation* – a term referring to the physiological and anatomical growth and development of cells (Gottlieb, 1998). The scope of predetermined epigenesis can be further expanded if it is assumed that structural maturation determines the function and capabilities of an individual, for example cell structures in neural networks are thought to govern cognitive function and behaviours (Battista et al., 2018), (Gottlieb, 1998).

The *nature* argument for development is based on predetermined epigenesis and can be summarised as the gene providing a blue print responsible for protein structure, cellular activity and organism function in the path illustrated in Figure 2.1 (Gottlieb, 1998). As species with similar gene pools display common developmental sequences, the essential role of genes in development is supported by Darwin's theory of evolution (Santrock, 2003). Meanwhile, explicit genes have been confirmed as responsible for specific developmental processes, for instance the homeobox gene accounts for the anatomical growth of complete segments in fruit flies (Schwartz, 1999).

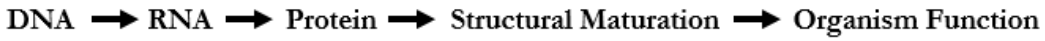


Figure 2.1: The biological role of genes (DNA) as the blue print that determines the functional capacities of organisms according to predetermined epigenesis. Image is adapted from Gottlieb (1998) pp.793

At the opposing pole the *nurture* argument suggests that the causal mechanism for development occurs through an individual's exposure to their social and cultural environment (Boyd et al., 2014). Psychologists such as Watson, Pavlov, and Skinner demonstrated that behaviour repetition and personality development could be predicted by exposure to environmental rewards and punishments (Santrock, 2003). Meanwhile some cognitive capabilities are clearly products of the surrounding environment; for instance, children's language is determined by imitating the words and sentences spoken to them, not because of an inbuilt mother tongue (Anisfeld, 2014).

The crucial place for nurture in development was confirmed by Tanner (1990) who identified differences in the height and cognitive performance of identical twins (sharing the same genome) who were raised in diverse environments. Tanner's (1990) evidence directly disputes

the role of the gene in human development as is understood from the perspective of predetermined epigenesis as it suggests the existence of feedback loops between gene, individual and environment (Gottlieb, 1998).

Probabilistic epigenesis is a contemporary theory of genetic expression suggesting the flow of information from gene to protein occurs as a set of bidirectional relations (Gottlieb, 1998). The central dogma might provide an incomplete theory for genetic expression as it does not account for evidence confirming the interaction between factors in the cell nucleus and cytoplasm: genes can be switched on or off by proteins (Tammen et al., 2013) and RNA (Dykxhoorn et al., 2003); RNA activity is affected by protein phosphorylation (Gottlieb, 1998); and a protein's conformational structure can be affected by other proteins (Koonin, 2012). In addition, Cirelli et al. (1996) identified the influence of extracellular neurotransmitters on intracellular genetic expression in rat brains, which could imply that probabilistic epigenesis extends beyond the cell.

According to Gottlieb (1998) genetic expression might be best understood by viewing the individual as a hierarchical system where bidirectional information passes between each level of increasing complexity from gene to chromosome, nucleus, cytoplasm, tissue, organism and finally environment. The hierarchical system might be further elaborated by considering Bronfenbrenner's (1977) ecological systems theory which ranks the environment according to its proximity to the individual from family (most proximal) to cultural (most distal).

The effects of both nature and nurture in human development could be accounted for by probabilistic epigenesis occurring within a hierarchical system. Bidirectional relations between each level of the system might explain how environmental stimuli influence biological processes to produce developmental change, as is illustrated in Figure 2.2. Rather than being a cut-off and predetermined blueprint, genes might be better viewed as followers (Pigliucci et al., 2010) *and* producers (Schwander and Leimar, 2011) in a larger and all-encompassing developmental process.

Probabilistic epigenesis informs contemporary psychological theory which defines lifespan development as the bidirectional *individual* ⟷ *context* relation (Lerner, 2006), (Lerner and Overton, 2010), (Fingerman et al., 2011), (Molenaar et al., 2003). It is now widely accepted that the mechanism driving human development is necessarily the interaction between nature *and* nurture. The *individual* ⟷ *context* relation places importance on all hierarchical levels from gene to the wider environment assuming each level is a mutually influential mechanisms driving change (Lerner and Overton, 2010).

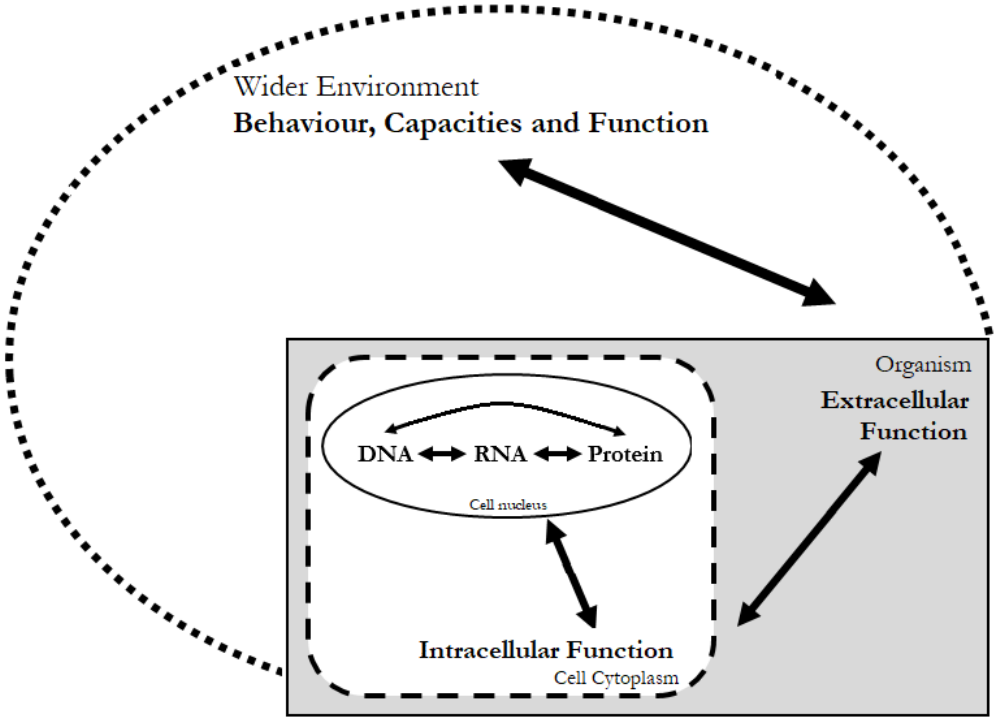


Figure 2.2: The biological role of genes (DNA) in development according to probabilistic epigenesis. Bidirectional relations exist between different hierarchical levels of the organism and its environment. Figure is adapted from Gottlieb (1998) pp. 797-798.

2.3.3 The Importance of Early Life

The most crucial and profound developmental changes occur during early life. Developmental plasticity refers to the extent that the development system is open to change at a specific point in time (Santrock, 2003). *Sensitive periods* are phases where development occurs most rapidly, and the influence of biological and environmental stimuli is unusually strong. It is thought that the earliest life stages are particularly plastic and most sensitive (Knudsen, 2004). For instance, the extent of change that occurs in the nine-months of prenatal development (i.e. transformative changes from cellular to living organism) is not comparable with nine months change that occurs in adolescence, (i.e. small changes in height, cognition etc.).

Developmental plasticity may be explained through the properties of neural circuits and transmission between neurons within these circuits (Knudsen, 2004). Connections between neurons in a circuit are themselves plastic meaning that the same stimuli can produce different responses at different times. The plasticity of neural circuits varies; for instance, neurons in the spinal cord always produce the same response where transmission occurs as reflexes which are predetermined by genetic expression; in contrast transmission between neurons in a brain

region called the amygdala is inconsistent and continuously adapts to environmental stimuli (Knudsen, 2004).

Knudsen (2004) describes three processes of neural plasticity that may explain the existence of sensitive periods. *Elaboration* is a process where connectivity between neurons is enhanced following repeated stimulation from sensory information – this reinforces existing connections and stimulates growth of new connections. In contrast, *elimination* is the removal of noise where neurons are pruned from the circuitry if they do not contribute to the response. Thirdly, Knudsen (2004) hypothesises *synapse consolidation* as a process where neurons with strong functional properties are bound together by cell adhesion molecules making them invulnerable to elimination.

The implication of developmental plasticity is that some abilities can only be learnt at specific periods in life; for example, young children are often able to learn a second language spoken without an accent, but this becomes increasingly difficult after the age of seven (Boyd et al., 2014). Equally, irreversible negative effects may occur if development is adversely affected during periods of high plasticity: prenatal development accounts for the organisation and formation of organs, limbs and the brain (Polin et al., 2016) and exposure to teratogens during this period have severe downstream effects including malformation, mental retardation and growth deficiencies (Gilbert-Barness, 2010).

As well as being a period associated with high levels of developmental plasticity, early life is important as it can have cascading effects continuing throughout later development. Early life is a time where fundamental skills are learnt which act as building blocks for the acquisition of future skills (Cunha and Heckman, 2007), (Santrock, 2003). Individuals with less proficient fundamental skills may find it difficult to develop abilities based on or requiring initial capacity. One example is to consider children with delayed early language acquisition who are likely to be at a disadvantage as they may be less able to develop other important skills such as communication, forming social relationships and learning through reading.

2.3.4 Lifespan Development as a target for Early Childhood Health Technologies

As discussed above, biological and environmental circumstances shape lifespan development and are likely to have the most profound impact if they occur early in life. Therefore, by targeting early life circumstances childhood health technologies could promote healthy development. Whilst certain biological circumstances are known to have severe adverse developmental effects, they are likely to prove difficult targets for health technologies. For example, genetic mutations are known to cause developmental disorders such as downs syndrome and autism (Weijerman and De Winter, 2010), (McRae et al., 2017) which could

theoretically be treated by targeting the malfunctioning genes. However, treatment would require a complex technology to deliver the new genetic material which might not be possible without impairing other important cellular mechanisms (Naldini, 2015).

Early environmental circumstances might provide a more feasible intervention target. There is abundant evidence linking harmful early environmental exposures to unfavourable developmental outcomes (Evans, 2006), (Ferguson et al., 2013) (Flora et al., 2012), (Walker et al., 2007). Several health technologies/social policies may have fostered lifespan development by reducing the exposure of children to damaging environmental conditions including: the banning of lead in manufacturing, a potent neurotoxin that can affect neurodevelopment, (Meyer et al., 2008); immunisation programmes that prevent the spread of infectious diseases such as bacterial meningitis associated with intelligence deficits and stunting (Segal and Pollard, 2004); and public education programmes that aim to reduce maternal consumption of nicotine and alcohol during pregnancy thus reducing the risk of developmental deformities and learning difficulties (Serdula et al., 1991), (Windsor et al., 2000).

It may also be possible for health and social policies to enhance child development by promoting positive environmental exposures. The introduction of universal schooling in low income countries has the potential to boost the development of a generation and provide long term national socioeconomic benefits (Glewwe, 2002). Meanwhile, cognitive and physical developmental benefits have been identified for Romanian orphans who were exposed to nurturing environmental circumstances after being placed in UK foster homes (Rutter, 1998).

2.3.5 The Lifetime Consequences of Development

Upon maturation, development provides an organism with the appropriate adaptive functions and behaviours that allow it to prosper within its environment (Lerner and Overton, 2010). As development is driven by the environment, the functional capacity and skills of modern humans are tailored towards the social and cultural world that they have been exposed to (Rogoff, 2003). The influential role of culture in driving change means that lifespan development influences a variety of outcomes relevant to health and social policy makers.

There is strong evidence establishing pathways between lifespan development and adulthood health. Theories linking health and development are based on the premise that health is the result of previous life events (e.g. accidents, exposure to infectious diseases, health behaviours etc.) and therefore the product of the individual ←→context relation. The idea of health as a product is central to Grossman's (1972) model of health demand where populations may produce health through investment; and Stoddart and Evans' (2017) health determinants framework where health occurs as a result of previous behaviour, biology, environment and

genetic factors. The role of development in producing health is the specific topic of the life course health development framework (Halfon et al., 2000), (Halfon et al., 2014) and the developmental origins of health and disease hypothesis (Hanson and Gluckman, 2008).

There is empirical evidence supporting the theoretical argument linking development to adulthood health: Malnutrition during early life influences the development of metabolic processes (Halfon et al., 2000) which increase the risk of hypertension, obesity, and diabetes later in life (Heindel and Vandenberg, 2015). Ben-Shlomo and Kuh (2002) summarise evidence suggesting sub-optimal development early in life increases risk of adulthood diseases such as cancer and bronchitis mediated through lifestyle factors. Meanwhile Lereya et al. (2015) find evidence that maltreatment and bullying can affect socioemotional development associated with poor adulthood mental health outcomes.

It is likely that the later effects of development are widespread and extend beyond health to affect outcomes in all areas of later life. A well-known pathway is the link between cognitive development and economic productivity, often mediated through the effects of cognition on academic achievement (Hanushek and Woessmann, 2008). Cunha and Heckman's (2008) technology of skill formation expands this pathway by formulising child development as a vector of cognitive *and* non-cognitive abilities that determine adulthood education, employment and crime outcomes.

Empirical evidence has also established the effects of development on a wide range of social outcomes for example: reduced cognitive function is correlated with adulthood unemployment, incarceration, crime, receipt of welfare and poverty (Cawley et al., 2001), (Herrnstein and Murray, 2010), (Frisell et al., 2012); and socioemotional development factors such as personality or behaviour are predictors of adulthood wage (Heckman, 2006), academic attainment (Hammer et al., 2017) and crime participation (Morizot and Kazemian, 2014), (Sampson and Laub, 2005).

2.3.6 Implications for the Indirect Estimation of Lifetime Effects

The literature review provides convincing theoretical and empirical evidence supporting the conclusion that there are potential lifetime effects associated with childhood health technologies which occur through a technology's influence on biological/environmental circumstances, and subsequently, the lifespan development process. This conclusion can be summarised as the effect pathway illustrated in Figure 2.3.

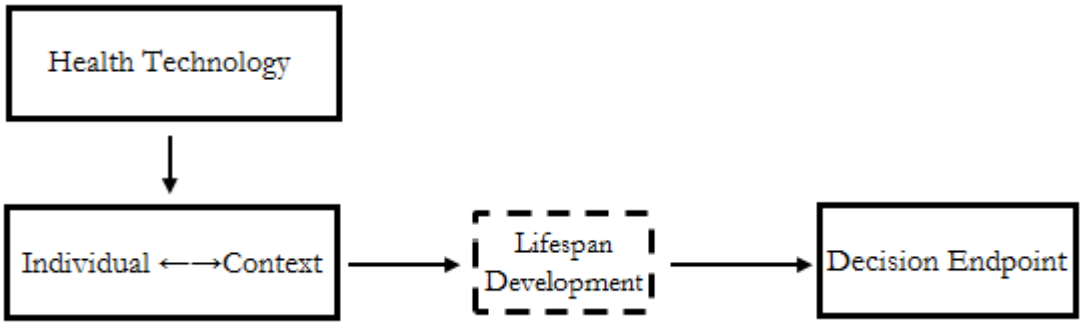


Figure 2.3: The pathway of effect between a childhood health technology and lifetime decision endpoints.

The identification of an effect pathway provides an important justification for the use of indirect estimation as a method to estimate the lifetime effects of child health technologies. The logical reasoning that underlies indirect estimation is to suggest that the effects of A (group allocation) onto C (decision endpoint) can be established by linking evidence regarding the effects of A onto B (intermediate endpoints) and the effects of B onto C. This reasoning *requires* an assumption that B occurs on an effect pathway that exists between A and C. According to the effect pathway in Figure 2.3, an appropriate intermediate endpoint that occurs between a health technology and a decision endpoint ought to be some measure of lifespan development. This method for indirect estimation is taken forward in chapter three which further discusses and estimates the effects of children exposed to symptoms of postnatal depression on intermediate outcomes that appropriately capture lifespan development.

2.3.7 Implications for Methods in the Economic Evaluation of Child Health Technologies

A further conclusion drawn from the lifespan development literature is to suggest that child health technologies may have fundamental differences compared with other health technologies in terms of the nature of their associated costs and benefits. Typically, the important costs and benefits associated with health technologies would be expected to be predominantly health related, whereas child health technologies affecting lifespan development may be associated with lifetime effects across a range of social outcomes including crime, education, employment, health and welfare. The size of cross sectoral lifetime effects may be substantial and important for decision makers, particularly if a technology is delivered during early life when development is highly plastic. The expected nature of lifetime costs and benefits could have implications when determining an appropriate philosophical framework and subsequent methodologies that

inform the economic evaluation of child health technologies – this forms the topic for discussion in the remainder of this chapter.

2.4 Philosophical Frameworks informing Economic Evaluation

2.4.1 Introduction

There are different philosophical frameworks informing the conduct of economic evaluation (Brouwer and Koopmanschap, 2000). Philosophical frameworks establish the methodological approach taken during analysis, determining the relevant outcomes (*decision endpoints*) on which to measure the costs and benefits of health technologies and designating appropriate methods to rank competing technologies (*decision rules*) (Brouwer and Koopmanschap, 2000). The decision to adopt a specific philosophy may be driven by the role assigned to economic evaluation in decision making. For example, economic evaluation might be viewed as a normative economics exercise meaning that explicit value judgements are made about preferred states of the world (i.e. what ought to be); here the role of analysis is to prescribe decisions by ranking technologies according to these states. In contrast, economic evaluation applying positive economics is concerned with observed states of the world (i.e. what is), where the role of analysis is to inform decisions according to the real-life requirements of decision makers (Claxton et al., 2007). The following section describes the competing philosophical frameworks of welfarism, extra-welfarism and the health/cross-sectoral decision maker's approach, and outlines the common approach adopted in the UK.

2.4.2 Welfarism

Economic evaluation can be conducted from the perspective of welfarism, a theory founded in principles of traditional welfare and normative economics (Brouwer et al., 2008). As described by Culyer (1991) welfarism is based on a fundamental principle which assumes that the social good is exclusively determined by the accumulation of individual *utility*. Within the welfarist framework, utility is a term representing the order of an individual's preferences over specific states of the world where higher levels of utility are achieved at higher ranked states (Brouwer et al., 2008). While utility is often informed by concepts such as happiness and well-being the terms are not synonymous, as utility *only* relates to individual preference (Brouwer et al., 2008).

A second principle of welfarism states that individual utility occurs as a function of the goods and services (commodities) consumed by the individual (Culyer, 1991). Healthcare might be viewed as such a commodity as consumption of healthcare can provide benefits (e.g. reduced pain, mobility, return to work etc.) which mean the post-consumption state is preferred to the

pre-consumption state. According to microeconomic theory, if viewed as an economic commodity in a perfect market, the efficient distribution of healthcare resources could be achieved through unregulated market forces, i.e. the invisible hand (Brazier et al., 2017). In this instance there would be no place for economic evaluation as efficient allocation of healthcare would be achieved through supply and demand.

However, the structure of publicly funded health systems does not represent a simple buyer-seller market place – the commodity of healthcare is associated with several interacting market failures including asymmetry of information, existence of externalities and lack of certainty (Brazier et al., 2017). From the perspective of welfarism the role of economic evaluation is to provide corrective evidence-based planned resource allocation that produces more efficient allocations of healthcare than would have been achieved in an unregulated and distorted market for healthcare (Brazier et al., 2017). To achieve efficient allocation of healthcare, welfarism attempts to establish whether the introduction of a health intervention would improve the social good (social welfare) or not. This means that the decision requires a valuation of society's preference (utility) towards the intervention (Brouwer et al., 2008).

A major challenge in welfare economics is the method used to aggregate individual utility to obtain an appropriate ordering of social welfare i.e. the identification of rules which consistently establish whether one state of the world is more *socially* desirable than another (Morris et al., 2007). Two contrasting views are; the utilitarian argument for efficiency where social welfare is assumed to be the sum of all individual utilities; and the Rawlsian argument for equity which values social welfare as the utility associated with society's least well-off member (Morris et al., 2007). Clearly efficiency and equity concerns are not always compatible, for instance a health intervention that increased utility in 10% of society's most well-off members and equally decreased utility in 5% of society's least well-off members would be accepted by utilitarians and rejected by Rawlsians.

In theory, society's views on efficiency and equity could be captured mathematically in what is termed a Bergson and Samuelson Social Welfare Function (Brouwer et al., 2008), (Feiwel, 2016). While capturing the ethical beliefs of society is likely to be a challenge, the major complication in defining a specific social welfare function is that utilities can only be reliably compared between individuals using ordinal (i.e. ordered preferences) and not cardinal (i.e. numerical) scales (Blackorby et al., 2002) (Hammond, 1991). Kenneth Arrow's impossibility theorem proved that individual ordinal preferences cannot be aggregated into rational social preferences unless under the restrictive conditions that a dictator's preference always dominates the preferences of any other individual in society (Feldman, 1989) (Hammond, 1991).

In the absence of an appropriate method to order social welfare, adoption decisions in economic evaluation could be informed by the Pareto principle: an intervention might only be recommended if it improved the utility of at least one individual without reducing the utility of *any* member of society (Johannesson, 1996). Whilst the Pareto principle is a weak value judgement likely to be acceptable to everyone, it is rarely realised for social interventions which are typically associated with opportunity costs. Kaldor and Hicks extended the Pareto principle into a more applicable compensation test framework suggesting that a change would be socially desirable if, hypothetically, those who experience an increase in utility following the change could provide adequate monetary compensation for individuals who were disadvantaged, whilst remaining better off themselves (Johannesson, 1996), (Morris et al., 2007).

Cost-benefit analysis (CBA) is a type of economic evaluation founded in welfarism which utilises the Kaldor-Hicks compensation test framework by monetising the costs and benefits associated with health interventions (Drummond et al., 2015), (Morris et al., 2007). In CBA, health interventions are recommended for adoption if they have a positive net benefit where the monetary value of social benefits achieved through consumption of the healthcare exceeds the monetary value of the associated costs (including opportunity costs) (Drummond et al., 2015).

Therefore, when conducting CBA, analysts must establish an appropriate way to monetise the costs and benefits associated with intervention. One method commonly applied to monetise outcomes in CBA is contingent valuation where hypothetical markets are constructed to establish an individual's willingness-to-pay (WTP) for a good. Contingent valuation using WTP is consistent with theoretical principles of welfarism as it can be interpreted in terms of compensation (Drummond et al., 2015).

Despite its theoretical grounding in welfarism there are some arguments regarding the accuracy and consistency of contingent valuation using WTP. For example: hypothetical WTP values derived from surveys may not be equivalent to the amount individuals would be willing to pay if actual markets existed (Kennedy, 2002); individuals might not be capable of providing an accurate valuation of healthcare products or health benefits (Donaldson et al., 2006); and, as individuals often view healthcare as a necessity, contingent valuation may lead to distributive issues as better-off individuals who are able to pay more are likely to state higher WTP values than less well-off individuals (Olsen and Donaldson, 1998).

In addition to issues faced when monetising costs and benefits, there are fundamental arguments against the use of compensation tests and monetary valuation in cost-benefit analysis. It may not be appropriate to establish a social net benefit through the summation of individual WTP. Blackorby and Donaldson (1990) claim that the summation of WTP requires an assumption and subsequent ethical judgement that income increases are equally socially

valuable no matter for whom they occur. It seems reasonable to question why a policy that, say, compensated the top 1% of earners by £X is necessarily of equal social value to a policy that compensated the bottom 1% of earners by same magnitude of £X. Furthermore, Boadway (1974) demonstrated methodological inconsistencies where the most desirable policies according to compensation test principles are not necessarily the best policies in terms of summed compensation variation (i.e. summed WTP).

There are further arguments against the application of welfare economics as a general framework to inform health economic evaluation. Welfarism is built on the assumption that individuals will make choices that rationally maximise their utility. It is often the case that choices are required when outcomes are uncertain. An individual, for example, might have to decide whether to have a surgery or not, knowing that the surgery has an associated probability of success and failure. In welfare economics, rational individuals would be expected to make decisions by multiplying the probabilities of the outcome occurring by the expected utilities and summing across all possible outcomes (Machina, 1987). However, Machina (1987) describes several choice scenarios where the assumption of utility maximisation under uncertainty fails to hold.

Meanwhile, policy recommendations from welfare economics are informed through the analysis of a perfect market in a "first best world". Real world economies are likely to be characterised by several market failures and imperfections including moral hazard, asymmetry of information and imperfect competition through oligopolies/monopolies. The application of first-best solutions to a second-best world is not necessarily appropriate and might reduce social welfare (Claxton et al., 2007), (Sculpher et al., 2005), (Richardson, 2007).

2.4.3 Extra Welfarism

Also founded within normative economic principles, *extra-welfarism* is an alternative perspective informing the conduct of economic evaluation which questions the fundamental assumption of welfarism. Arguing against the exclusive use of individual utility to determine the social good, Sen (2008) suggests that utility occurs through an individual's mental reaction to a good. Sen provides an example of a deprived individual who may generate higher levels of utility from smaller levels of consumption when compared with a less deprived individual, as they have learnt realistic desires and may "take pleasure in small mercies". Goods are likely to have benefits beyond utility in terms of what they enable an individual to do or be (Sen, 2008). Welfarism might be flawed given its failure to account for objectively desirable social benefits that occur beyond individual utility maximisation.

Extra-welfarism, as termed by Culyer (1991) and applied in health economics, is an alternative philosophical perspective for economic evaluation which suggests that additional factors should be included as an objective to maximise alongside individual utility. One such factor is health which is likely to be an important consideration when informing decisions regarding the allocation of health interventions (Brouwer and Koopmanschap, 2000), (Brouwer et al., 2008).

Morris et al. (2007) suggest health may be accounted for within economic evaluation by considering the effect of intervention on the quality of an individual's health state (morbidity) and their expected length of life (mortality). An example of such a measure is the Quality Adjusted Life Year (QALY) which combines multiple concepts of morbidity and mortality into a single index (Drummond et al., 2015). QALYs are constructed using questionnaires which rank patients into one of many specific health states and use methods such as the standard gamble or time trade off to establish society's preference for each health state. The preference-based valuation produces cardinal values termed health-related quality of life (HRQoL) and these are accumulated over the remainder of an individual's life expectancy to derive their lifetime QALYs (Drummond et al., 2015).

Cost-utility analysis (CUA) is partially consistent with the principles of extra-welfarism as it is a specific type of economic evaluation establishing the net-benefit of an intervention by comparing the ratio of incremental costs with its incremental benefits measured as QALYs (Drummond et al., 2015). A new health intervention is recommended for adoption if the *incremental* cost-effectiveness ratio (ICER) is below a pre-specified threshold value when compared with current practice (Drummond et al., 2015). By comparing ICERs to a cost-effectiveness threshold, which is defined as the 'cost per unit of health benefit forgone' (Woods et al., 2016), CUA accounts for the opportunity costs that occur through the decision not to adopt a competing intervention.

Brouwer et al. (2008) suggest the extra-welfarist perspective might provide a more pragmatic approach to economic evaluation not requiring adherence to restrictive Pareto principles or unreliable methods of valuation associated with contingent valuation in welfarism and CBA. Explicit between-person comparisons can be obtained on measures of health, such as the QALY, that cannot be obtained for ordinal utilities. As stated above, the QALY is a cardinal measure based on social preferences meaning QALYs are implicitly assumed to be of equal social value irrespective of who achieves them (i.e. a QALY should always equal a QALY) (Whitehead and Ali, 2010).

There is, however, some debate regarding the consistency of CUA with the principles of extra-welfarism. The objective of extra-welfarism is not exclusively health driven but is the maximisation of the social good of which health is only one factor alongside many others

including individual utility. For CUA to be considered as fully consistent with the "extra" welfarist paradigm, QALY maximisation would need to capture individual utility maximisation. Dolan and Edlin (2002) describe fundamental theoretical restrictions that may prevent this, for instance, a health intervention might increase individual utility by improving health *and* capacity to work. It is debateable whether both arguments are captured when measuring QALYs. Acknowledging the limitations of the QALY, Coast et al. (2008) argue that the health centred perspective of extra-welfarism that informs CUA focuses on function rather than "freedoms (capabilities) to pursue health improvement".

Wider measures of well-being have been developed for economic evaluation and are beginning to be applied in some empirical literature: A well-being QALY has been advocated by Cookson et al. (2016) which attempts to operationalise Sen's capabilities' approach by accounting for health as well as other factors related to income and the consumption of goods. The Investigating Choice Experiences for the Capabilities in Adults (ICECAP-A) is a preference based capability measure that has been applied in UK populations by Mitchell et al. (2017) and contains items for stability, attachment, autonomy and achievement; meanwhile, the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) is a non-preference based measure used to assess mental well-being where participants score 14 items related to positive mental health including confidence, optimism, relaxation and cheerfulness (Tennant et al., 2007).

The different dimensions used in Cookson's well-being QALY, the ICECAP-A and the WEMWBS however demonstrate a pragmatic limitation faced in the application of well-being measures: there is a general lack of consensus about what the key theoretical components of well-being are, and how these ought to be measured (Cookson, 2005; Nussbaum, 2003). This might explain why, despite being theoretically inconsistent with some aspects of extra-welfarism, the health centric QALYs is often the preferred outcome measure in CUA.

2.4.4 A Health Centric Decision Maker's Approach

Both welfarism and extra-welfarism are structurally similar in the sense that they provide a theoretical basis for healthcare allocation decision making within a normative economic framework. A contrasting decision maker's approach (DMA) views economic evaluation within a positive economics framework and aims to inform rather than prescribe decisions (Brouwer and Koopmanschap, 2000). The DMA addresses pragmatic issues faced by policy makers and moves away from questions of what ought to be i.e. maximising the social good, and towards questions of what is, i.e. achieving the explicit objectives of a *legitimate* social decision maker (Brouwer and Koopmanschap, 2000).

The pragmatic issues faced by decision makers in the UK often relate to the appropriate distribution of collective funds which are allocated into sector specific budgets providing diverse public services such as health, education, transport, crime prevention etc. If the decision maker's responsibility is limited to a specific sector then their role might be summarised as a constrained optimisation problem, where they wish to maximise some measure of benefit within that sector subject to an exogenous budget constraint (Sculpher et al., 2005).

The appropriate methods for economic evaluation for the DMA are determined by the specific optimisation problem faced. In general, cost-effectiveness analysis (CEA) provides a tool for non-specific constrained optimisation problems where the benefit of an intervention can be obtained on the primary outcome of interest to the decision maker (Drummond et al., 2015). For example, economic evaluations conducted from a DMA and informing policy decisions in the crime sector might value the benefits of intervention in terms of reduced crime/incarceration rates, whereas decision makers in the education sector might assess benefits on endpoints including graduation success or scholastic achievement.

The National Institute for Health and Care Excellence (NICE) is an institution which might be thought of as a legitimate decision maker in the health sector as it has delegated authority to make healthcare policy recommendations from a democratically elected central government (Sculpher et al., 2005). The role of NICE (2013) is reflected in their perspective for economic evaluation which is to maximise health benefits given the exogenous resources made available to the National Health Service (NHS) and the Personal Social Services (PSS). The methodology proposed by NICE (2013) is to conduct a CUA using QALYs and healthcare costs incurred by the NHS and PSS as decision endpoints.

There are considerable cross-overs in the methodology for economic evaluation as prescribed by a health centric DMA and the extra-welfarist perspective and this may have led commentators such as Brouwer et al. (2008) to term the DMA as a "branch" of extra-welfarism. This terminology is avoided here given the fundamental differences in normative and positive economic principles informing each perspective. From the health centric DMA, cost-utility analysis might be best thought of as a specific form of cost effectiveness analysis using a group of common methodologies to inform adoption decision in the health sector (Drummond et al., 2015). Unlike the extra-welfarist perspective, the sole objective of the health centric DMA is to achieve as much health as possible and therefore does not require measures to capture both health and utility. It follows that health centric measures like the QALY are most suitable when establishing intervention benefits for a health centric DMA perspective when compared with more generalised measures of capability/well-being like the ICECAP-A/WEMWBS.

According to Claxton et al. (2010), the health centric DMA, which maximises health given a budget for healthcare, is only a reasonable option if the following five criteria are met: (i) the social objective is to improve health; (ii) measures of health gained and forgone capture enough aspects of the social objective; (iii) the budget is exogenous (i.e. cannot be changed by the decision maker); (iv) the costs associated with the health technology are all in terms of health forgone; and (v) all the benefits of the intervention are in terms of health.

Claxton et al. (2010) justifies a health centred social objective *if* the budget for health is a legitimate expression of social value. In this case maximising measures of health (QALYs) subject to the healthcare budget in cost-effectiveness analysis might be entirely appropriate as society's willingness to pay for health is represented by the money made available in the healthcare budget. The shadow price of the budget (society's willingness to pay) is accounted for using a cost-effectiveness threshold which compares health gained from new interventions with health forgone in displaced interventions as is indicated in the decision rule in Equation 2.1 (Claxton et al., 2007), (Claxton et al., 2010).

Equation 2.1: Decision Rule from a Health Centric Decision Makers Perspective

Decision rule: Adopt an intervention if;

$$[\Delta h - \Delta C_h/k] > 0$$

Where:
Δh= incremental health effects e.g. QALYs
ΔC_h= incremental costs that fall on the healthcare budget
k= The cost effectiveness threshold (the additional costs that would displace 1 unit of health elsewhere in the healthcare system)

Notes: Taken from Claxton et al. (2010). The incremental costs and benefits of an intervention are identified versus a relevant competitor i.e. usual care or an alternative intervention strategy.

2.4.5 The Cross-Sectoral Decision Maker's Approach

2.4.5.1 Structure of Decision Problem

The final two above criteria specified by Claxton et al. (2010) are unlikely to hold for all healthcare interventions which may have spill-over effects in other sectors, for example; health interventions often incur costs and/or benefits for carers and may also have effects on economic productivity allowing sick individuals to return to work. The health centric DMA may be defensible if these spill-over costs and benefits are inconsequential to the overall decision (Drummond et al., 2015). However, in some cases cross-sectoral effects are substantial, such as in the evaluation of public health policies, and it might not be advisable to ignore these outcomes as this may lead to bias during decision making (Claxton et al., 2010).

To account for spill-over effects, the health centric DMA can theoretically be extended to a *cross-sectoral DMA* with an objective function that maximises health with respect to healthcare costs, but also maximises other relevant outcomes across different sectors (e.g. education, crime, transport) with respect to costs incurred by that sector.

According to Claxton et al. (2007) the more complicated optimisation problem with multiple measures of benefit and multiple costs cannot be solved mathematically given the information currently available to analysts. Considering this, Claxton et al. (2007) suggests a theoretical rule for cross-sectoral decision making which employs compensation test principles; for instance, if an intervention had a small negative net health benefit but a large positive net education benefit, it could be adopted if the education sector adequately compensated the health sector. However, this method for decision making cannot be applied as it requires administrative procedures that transfer actual monetary compensation between sectors which are not currently available (Claxton et al., 2007).

Claxton et al. (2010) conducted further research and established a conceptual framework consistent with the cross-sectoral DMA that can account for costs and benefits occurring in multiple sectors. Within this framework Claxton et al. (2010) assume that budgets are sub-optimal expressions of social value (i.e. do not entirely represent society's willingness-to-pay). This means that consumption in one sector does not necessarily equate to the same level of consumption in another sector (i.e. £1 spent in health is *not* necessarily equivalent to £1 spent in education). Claxton et al. (2010) suggest that some other legitimate valuation of sector-specific outcomes relative to consumption is required and formulised this by including a consumption value for health *(v)* in the decision rule described in Equation 2.2.

Equation 2.2: Decision Rule from a Cross-Sectoral Decision Makers Perspective

Decision rule: Intervention should be adopted if;

$$v.[\Delta h - \Delta C_h/k] - \Delta C_c > 0$$

Where:

v=	The consumption value of heath (the amount of consumption in the wider economy regarded as equivalent to 1 unit of health)
ΔC_c=	The net effects which do not fall on the healthcare budget expressed as net consumption cost to the wide economy.
Δh=	Incremental health effects
ΔC_h=	Incremental costs that fall on the healthcare budget
k=	The cost effectiveness threshold

Notes: Taken from Claxton et al. (2010). The incremental costs and benefits of an intervention are identified versus a relevant competitor e.g. usual care or an alternative intervention strategy.

The consumption value for health provides a mechanism for weighting costs and benefits falling on the wider economy. Claxton et al. (2010) suggest this value may fall between zero and one placing it between the two policy extremes that either (a) ignore cross-sector costs and benefits or (b) treat cross-sector costs and benefits as if they fall directly on the healthcare budget. Whilst the precise value of this weighting is not known the cross-sectoral DMA could operationalise Equation 2.2 in an economic evaluation by providing net-benefit calculation across a range of feasible consumption values for health. This information would be useful for decision makers as it would illuminate the potential biases that may occur when adopting a health centric DMA which excluded cross-sectoral costs and benefits.

2.4.5.2 Measuring Intervention Benefits

A further consideration if adopting a cross-sectoral DMA is establishment of an appropriate method to assess intervention benefit. Intuitively, the extension of the decision perspective beyond health might appear to favour extended measures of capability/well-being (e.g. ICECAP-A/WEMWBS) over narrower and more health centric measures (e.g. QALYs). However, the optimisation problem specified by the cross-sectoral DMA assumes budgets are assigned according to specific objectives. For example, the health budget is delivered to improve population health, the education budget to improve education, the transport budget to improve transport etc. It is not likely that sector specific objectives are appropriately captured by dimensions like enjoyment, achievement, and stability within extended well-being QALYs.

In theory, researchers applying a cross-sectoral DMA might adopt *multiple* sector specific outcomes measures. As with the health centric DMA, QALYs provide a measure of intervention benefit directly relevant to the objectives of the health sector. In addition, the cross sectoral DMA might include QALY equivalent measures specific to the requirements of other sectors e.g. education QALYs, crime QALYs etc. One such example is the Adult Social Care Outcomes Toolkit (ASCOT) a measure of *social care* related quality of life (Peasgood et al., 2014).

There are, however, two issues regarding the use of multiple sector specific measures of benefit with the cross-sectoral DMA. Firstly, intervention benefits might be double counted if outcome measures have overlapping items – a problem that is illustrated by QALY and ASCOT measures which both contain items assessing psychological function (Peasgood et al., 2014). Secondly, and more fundamentally, appropriate QALY equivalent outcome measures have yet to be developed within each sector. Economic evaluations in the education sector, for example, do not typically adopt specific outcome measures and instead measure education benefits in terms of monetary returns accrued through increased earnings, akin to willingness to pay methods in cost-benefit analysis (Cookson et al., 2016; Levin and Belfield, 2015).

In the absence of relevant and unique sector specific outcome measures, a pragmatic solution to assigning benefits within the cross sectoral DMA could (i) assign QALYs as the outcome measure in the health sector, and (ii) assign monetary returns as benefits specific to each of the other sectors (i.e. WTP). Monetary returns and QALYs could be combined within a single decision rule by monetising QALYs using the cost-effectiveness threshold (*k)* and weighting monetary benefits using sector specific consumption values relative to health (*v)*. An example of the adapted and expanded decision rule including benefits and costs for health, education and crime sectors is reported in Equation 2.3. The equation is consistent with those described by Claxton et al. (2010) who categorised decision endpoints outside of the health sector in monetary terms as either: (i) direct costs of treatment that are not incurred by the health sector or (ii) indirect effects on the wider economy that are external to patients and their family.

Equation 2.3: Decision Rule from a Cross-Sectoral DMA in Health, Education & Crime

Decision rule: Intervention should be adopted if;

$$[\Delta h - \Delta C_h/k] + [\Delta R_e - \Delta C_e]/v_e + [\Delta R_c - \Delta C_c]/v_c > 0$$

Where:

$\Delta h =$	Incremental QALYs
$\Delta C_h =$	Incremental costs that fall on the healthcare budget
$k =$	The cost effectiveness threshold
$\Delta R_e =$	Incremental monetary returns for the education budget
$\Delta C_e =$	Incremental costs that fall on the education budget
$v_e =$	The consumption value of heath relative to education (the amount of consumption in education regarded as equivalent to 1 unit of health)
$\Delta R_c =$	Incremental monetary returns for the crime budget
$\Delta C_c =$	Incremental costs that fall on the crime budget
$v_c =$	The consumption value of heath relative to crime (the amount of consumption in crime regarded as equivalent to 1 unit of health)

Notes: Equation is adapted from those presented by Claxton et al. (2010). The incremental costs and benefits of an intervention are identified versus a relevant competitor e.g. usual care or an alternative intervention strategy.

2.4.6 Economic Evaluation in the UK

2.4.6.1 UK Central Government

The publicly funded UK health system operates at many different levels. Healthcare provision is influenced by national policies made by the government's Department of Health through to individual choices made by GPs/consultants regarding the healthcare available to specific patients (Corbacho and Pinto-Prades, 2012). Whilst the UK municipalities (local authorities) are given some autonomy regarding service provision, over recent decades, the UK

government has taken a central role in decision making and funding for healthcare and other public services including education, criminal justice and transportation (Walshe et al., 2016).

In 1999, the Department of Health began delegating national (central) decision-making authority and NICE was set up with the objective of providing recommendations on the adoption of new and existing health interventions, (Buxton, 2006). Currently, in England the NHS has a legal constitution which specifies that patients have the right to treatments identified as cost-effective in NICE technology appraisals if GPs or consultants deem them to be clinically appropriate (Raftery, 2014). Several other bodies now exist across the UK whose role is also to prescribe or advise healthcare decisions including: the Scottish Medicines Consortium, the Scottish Intercollegiate Guidelines Network, the All Wales Medicines Strategy Group, the UK National Screening Committee, and the Joint Committee on Vaccinations and Immunisations (Buxton, 2006), (Corbacho and Pinto-Prades, 2012), (Raftery, 2014).

Economic evaluation plays a fundamental role in informing the decision-making process undertaken by NICE (and the other UK healthcare jurisdictions) and this can include internal primary analyses or analyses submitted by external (often academic) institutions (Buxton, 2006), (Raftery, 2014). In fact, NICE has been a primary driver in the development of methods for economic evaluation and has been influential in promoting economic evaluation as a method to facilitate healthcare decisions across Europe (Corbacho and Pinto-Prades, 2012). To inform the methods for economic evaluation NICE has published a reference case (2013) which all external submissions are required to conform to if they are to be considered in NICE policy appraisals.

The NICE (2013) reference case adopts a health centric DMA as the primary perspective of analysis. As stated in the NICE (2013) reference case, the primary submission should be a cost-utility analysis that considers the health benefits of interventions measured as QALYs and healthcare costs incurred by the NHS and Personal Social Services (PSS). The health centric perspective adopted by NICE reflects both the assumed political role of the UK central government (i.e. to provide budgetary funding for specific public services including health) and the role assigned to NICE as an authority for healthcare decision making. In addition, NICE (2013) do consider secondary "non-reference case" submissions if researchers are able to justify theoretically why the reference case methods are inappropriate. In these cases, it would be appropriate for analysts to conduct a primary reference case analysis which is presented alongside any additional evaluations employing non-reference case methods.

NICE (2013) argue that the reference case provides a "crucial" approach for healthcare decision-making as it provides methodology that can be applied consistently and transparently across a wide range of healthcare decisions in the UK. Consequently, analysts may prefer to

adopt NICE (2013) reference case methods regardless of their own view on the appropriate perspective for analysis as these methods may provide the most likely route to policy influence in the UK.

2.4.6.2 Devolution and Service Integration

Section 2.4.6.1 described the role of modern UK governments, which have tended to operate and fund health (and other public services) at a central level. Recently, some political initiatives have reversed the trend towards centralisation by transferring power to more local political structures through *devolution*. Examples include deals for Scotland, Northern Ireland and Wales in 1999; regional governments in England in 2004; and a £6 billion per year agreement for Greater Manchester known as "Devo Manc" in 2015 (Walshe et al., 2016).

According to Walshe et al. (2016), a benefit of devolution is the ability to integrate public services as smaller local governments are better placed to deliver collective care across multiple organisations, when compared with central based systems that are disconnected through complex governance structures. A primary objective of the Devo Manc project, for example, has been the integration health and social care services which now fall under a single budget (Gains, 2015).

The integration of budgets under devolution has implications for the appropriate perspectives from which to conduct economic evaluation. The health centric DMA does not appear to sufficiently fulfil the requirements of, say, a Devo Manc decision maker whose objective is likely to be the maximisation of both health *and* social care subject to an integrated budget for health *and* social care. Theoretically, it might be desirable to extend the health centric perspective to a health *and* social care centric DMA requiring analyses to consider all costs falling on the integrated budget, and possibly measuring intervention benefits on an expanded QALY which includes additional items relevant to social care.

Pragmatically, tailoring perspectives for economic evaluation to the exact requirements of devolved local regions/authorities might be unrealistic: Regions may differ in terms of how and which services/budgets are integrated, and therefore different regions would require a unique analytical perspective and subsequent set of methods relevant to their specific political organisation. The adoption of a cross-sectoral DMA might be more feasible approach as it: (i) accounts for monetary costs and benefits falling in multiple sectors and is therefore likely to be consistent with the structure and objectives of most devolved governments; and (ii) would be applicable for analyses across different devolved regions thus maintaining a consistent set of national methodological standards.

Whilst there are no specific guidelines on methods for economic evaluation in light of recent devolution deals, the precedent for moving towards a cross-sectoral approach may have been set by NICE with respect to their guidelines for analyses of public health interventions. Funding for public health in England has been devolved to local authorities (Greer, 2016); such interventions are known to have costs and consequences that span multiple sectors and budgets (Claxton et al., 2010), and most NICE guidelines for economic evaluations of public health interventions include costs across multiple sectors beyond health (Hinde et al., 2017). *If* future devolution initiatives continue to integrate budgets for public services, it may be that a cross-sectoral (rather than health centric) DMA becomes the primary analytical perspective recommended by NICE in the UK.

2.4.7 Conclusion

The review of the health economics literature above summarises the key perspectives informing the conduct of economic evaluation, identifying pragmatic and theoretical arguments for and against the perspectives of welfarism, extra-welfarism and the health/cross-sectoral DMA. Section 2.5 draw on this background and consider the selection of methods and evaluation vehicle to be used in the applied economic evaluation which is the subject of the empirical research in this thesis. The discussion provides answers to the questions posed in 2.1 about the *type* of economic evaluation to be conducted, the decision rules to be applied and the decision endpoints to be used.

2.5 Synthesis and Future Research Methods

2.5.1 Perspectives informing the Economic Evaluation of Early Childhood Health Technologies

In the UK, decision making is influenced by institutions such as NICE and methods for health economic evaluations generally follow those prescribed by NICE which requires primary submissions to adhere to methodologies reported in the reference case (NICE, 2013). As seen above this specifies that economic evaluations adopt the perspective of a health centric DMA and conduct a CUA where the benefits of interventions are measured in terms of QALYs and costs are limited to those incurred by the NHS and PSS. NICE may also consider secondary analyses alongside reference case analyses if theoretical justification is provided that identifies why the reference case methods may be insufficient. If UK healthcare decision making requires consistent methods to be applied for all allocative decisions, then the economic evaluation of

UK child health interventions should follow NICE (2013) guidelines and adopt the health centric DMA as a perspective for analysis.

Based on these conclusions, the applied economic evaluation conducted in this thesis assesses the cost-effectiveness of screening strategies for postnatal depression by establishing the costs and benefits of each strategy using children's lifetime QALYs and direct healthcare costs as *decision endpoints*. The analysis identifies the cost-effective strategy by applying the *decision rule* described by Claxton et al. (2010) and reported in Equation 2.1.

The review of the lifespan development literature identified the potential effects of early life diseases/interventions on child development and revealed the influence that development has on all manner of later life outcomes. Therefore, and like the public health literature, early childhood intervention may be associated with spill-over effects in sectors beyond health that are consequential in decision making. As was identified by Claxton et al. (2010), reference case analyses which exclude important cross-sectoral effects could result in biased decisions being made. For instance, if, in general, early life interventions that foster development produce larger cross-sectoral benefits than health benefits, then the exclusion of these benefits might result in decisions that do not recommend early interventions despite them providing a social benefit.

The potential for substantial cross-sectoral benefits and the subsequent possibility of bias provides a theoretical reason to diverge from the reference case methods. Therefore, based on the findings in this literature review, it is suggested that the economic evaluation of child health technologies in the UK provide *a second analysis* alongside the reference case analysis. The second analysis should extend the perspective to that of a cross-sectoral DMA and consider the costs and benefits associated with health interventions that occur in sectors outside of health.

The *decision rule* suggested by Claxton et al. (2010) for the analysis of public health interventions (Equation 2.2) and the expanded Equation 2.3 provide a conceptual framework for analysing the social costs of interventions when adopting a cross sectoral DMA. The *decision endpoints* relevant to the cross-sectoral DMA perspective include QALYs and direct healthcare costs, but also require the costs and benefits of interventions to be assessed on outcomes in sectors outside of health. In theory, cross-sectoral effects could be captured as costs incurred by other budgets and on QALY equivalent outcomes that measure benefits specific to each sector i.e. education QALYs, crime QALYs etc. However, no QALY equivalent exists in these sectors where public policy is usually evaluated in a Cost-Benefit Analysis (Cookson et al., 2016). Consequently, the spill-over effects of early intervention might be best measured in monetary terms as lifetime costs and savings incurred by sector specific budgets.

More research is required to establish the precise consumption values for health with relation to consumption in other sectors. However, a set of secondary analyses could provide useful

information to decision makers by investigating cost-effectiveness across a range of feasible consumption values, illustrating how different consumption values might affect decisions regarding cost-effectiveness.

It is debatable which of the analyses ought to be presented to decision makers as the primary/secondary results. On one hand it is desirable to include all important costs and benefits associated with health technologies, but on the other hand it is equally desirable to uphold consistent methods for economic evaluation across all health allocation decisions. By adopting the perspective of a health/cross-sectoral decision maker the role of economic evaluation is assumed to inform rather than prescribe decisions. Therefore, the choice regarding the primary/secondary analysis becomes less important as decision makers can select the most appropriate analysis given their objectives *if* the analyses clearly state the philosophical frameworks that inform the methodology.

To remain consistent with NICE (2013) guidelines this research adopts the frameworks of healthcare DMA for the primary analysis, and the cross-sectoral DMA for the secondary analysis.

2.5.2 Selecting an Appropriate Vehicle for Economic Evaluation

The final consideration in this chapter is to briefly discuss the appropriate analytical technique and use of empirical evidence within economic evaluation consistent with the selected perspective. As described by Sculpher et al. (2006) economic evaluation can be conducted using one of two *vehicles for analysis*. Firstly, trial based economic evaluations are an expansion of the traditional randomised clinical trials(RCTs): participants are randomly allocated into intervention/control groups; follow up information is collected related to each participants healthcare usage and health utility; and mean incremental healthcare costs and QALYs (or equivalent measure of health/clinical benefits) are used to calculate ICERs between the intervention/control groups based on the observed data (Sculpher et al., 2006).

According to Sculpher et al. (2006) trial based economic evaluations should not be used as the sole vehicle for economic evaluation for the following pragmatic restrictions: it is not possible to compare all the relevant treatment strategies within single trials; RCTs often fail to account for the longer term (lifetime) effects of intervention as they usually have truncated time horizons due to high costs; the results of trials are conducted in specific populations that might not be generalisable to the population being targeted by the policy; trials are not able to measure all relevant evidence and cannot incorporate evidence from outside sources; and finally, trial-based evaluation cannot establish the probability of the correct decisions being made and therefore cannot quantify the opportunity costs of making a wrong decision.

For the economic evaluation of early interventions, a trial-based approach appears particularly insufficient given the expected duration of treatment effects, which could occur decades after treatment initiation. The inadequacy of a trial-based approach is exemplified by the applied economic evaluation that is the subject of this thesis where it was not feasible to conduct an RCT which allocated mothers into the relevant treatment groups (different screening strategies for postnatal depression) whilst identifying lifetime treatment effects for children on the relevant *lifetime* decision endpoints.

The second vehicle for economic evaluation described by Sculpher et al. (2006) is the decision analytic model (DAM), a mathematical model which conceptualises disease progression. Usually, a hypothetical patient population is simulated through the different disease states within a DAM. Analysts split the hypothetical population into equal groups which receive all relevant competing intervention strategies. Each intervention strategy differently affects the probability of progression between disease states. Healthcare costs and QALYs are assigned to each disease state and the cost-effective strategy is identified by summing costs and QALYs for (and identifying ICERs between) the hypothetical treatment groups (Briggs et al., 2006).

The parameters within DAMs (i.e. probabilities of transmission between disease/remission states, healthcare costs and QALYs) are informed through secondary data usually relying on evidence obtained from randomised clinical trials or observational studies but can also potentially include evidence from less robust sources such as expert elicitation (Sculpher et al., 2006). The use of secondary data lifts the pragmatic restrictions faced by trial-based evaluations, allowing DAMs to compare all treatment options over the appropriate time horizon in populations relevant to the decision maker (Sculpher et al., 2006).

In addition, DAMs can account for all available evidence by applying a bayesian approach to statistical analysis which views probabilities as subjective beliefs updated based on new information (Cooper et al., 2004). This is opposed to the traditional frequentist view that assigns probability through observations made from repetitive experiments (Cooper et al., 2004). The application of Bayesian techniques allows economic evaluation to be conducted in a way that mirrors a real-life decision-making process: initial decisions are made, new evidence is gathered, and decisions are revised given the new evidence (Claxton et al., 2007). Using mathematical simulation techniques, bayesian DAMs allow probabilities of cost-effectiveness to be calculated alongside mean estimates of cost-effectiveness which is an important consideration in decision making (Claxton, 1999).

It is the view of Sculpher et al. (2006) that DAMs provide the only vehicle for analysis that can meet all the objectives of the decision maker. When compared with trial-based evaluations, DAMs are better suited to deal with some empirical challenges faced when conducting

economic evaluation of interventions that affect child development. For example, there is no requirement (as needed for trial-based evaluations) to directly compare all treatment strategies within a single study over the full life cycle. Instead the structure of the DAM specifies the evidence that is required which can be obtained and synthesised from multiple sources e.g. RCTs, observational studies, expert elicitation etc.

2.6 Summary of Conclusions and Empirical Research Design

2.6.1 Conclusions

By bringing together evidence from the lifespan development and the economic evaluation literature this chapter identifies several methodological approaches informing the primary research in this thesis.

Firstly, the identification of an effect pathway suggests a design for indirect estimation where the effects of early life circumstances on lifetime decision endpoints can be estimated by: (i) observing effects on intermediate endpoints measuring lifespan development, and (ii) extrapolating intermediate development measures to effects on decision endpoints across the lifespan.

Secondly the review suggests an appropriate perspective and methodology for conducting economic evaluation of child health technologies. To provide the most influence on UK policy, the design of the economic evaluation in this thesis is consistent with the NICE (2013) reference case i.e. a cost-utility analysis that identifies benefits of health technologies on QALYs and costs in terms of those borne by the healthcare budget.

As early technologies affecting child development are likely to be associated with substantial cross-sectoral effects, a second analysis is included in addition to the reference case analysis. The non-reference case analysis applies a cross-sectoral decision maker's approach establishing effects of interventions using QALYs, healthcare costs, costs and cost savings in sectors outside of health (including economic productivity) as decision endpoints. The second analysis includes a feasible range of consumption values for health.

Bayesian DAMs provide the most appropriate vehicle for this economic evaluation. The models are particularly useful given some of the methodological difficulties likely to be faced when estimating the lifetime effects of child health technologies as the structure of the model specifies the exact research priorities, and these can be informed by evidence from multiple and diverse sources.

2.6.2 Research Design for the Applied Economic Evaluation

The applied economic evaluation in the thesis adopts the perspectives, methodologies and analytical vehicle suggested in this chapter. The economic evaluation assesses the cost-effectiveness of screening for postnatal depression, extending an existing cost-effectiveness study by NICE (2018) which was limited as it did not formally consider the costs and consequences of screening on children's lifetime decision endpoints.

The economic evaluation in chapter six includes children's lifetime effects by adding additional states for children exposed to symptoms of postnatal depression. The underlying structure of the DAM follows the NICE (2018) design: screening is administered 6 weeks after childbirth which results in women being assigned to one of four diagnosis groups (i.e. true positives, true negatives, false positives false negatives); mothers then receive treatment according to their diagnosis group; and the model culminates by grouping women into either depressed or non-depressed health states. Figure 2.4 depicts the treatment part of the NICE (2018) DAM for mothers who are correctly (true positives), and wrongly (false positives) diagnosed with depression by a screen. The full structure of the DAM is detailed later in chapter six.

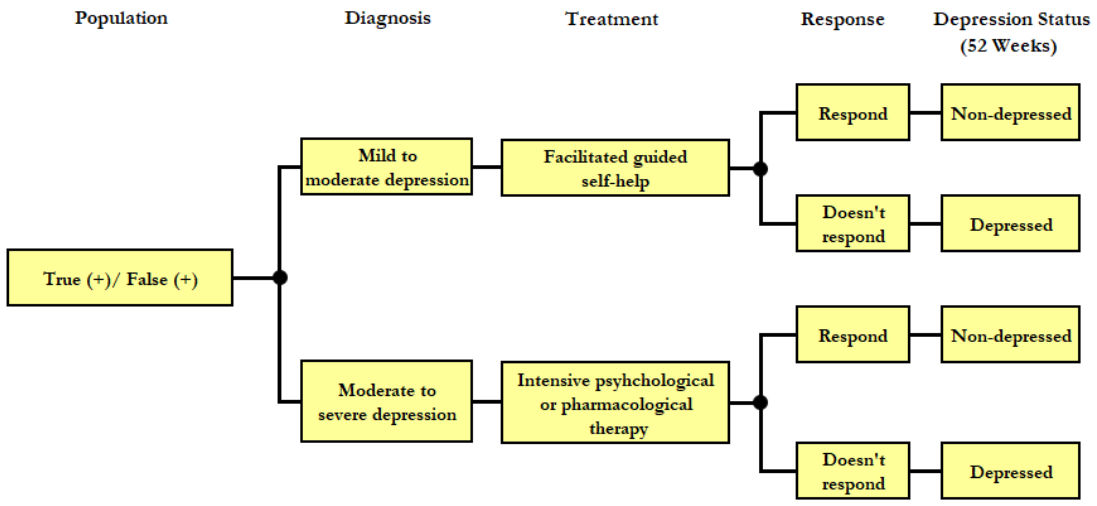


Figure 2.4*:* Depicts the true positive and false positive treatment components of the decision tree used to assess the cost-effectiveness of screening for postnatal depression by NICE (2018).

The DAM in this thesis adds health states for children who are either assumed to be exposed or not exposed to symptoms of maternal depression according to their mother's depression status at the culmination of the DAM. The addition of an exposure state for children is illustrated in Figure 2.5. Informed through the structure of the DAM, the lifetime effects evidence needed

for the application in this thesis requires the estimation of the incremental effects in children exposed to postnatal depression versus children who are not exposed to postnatal depression.

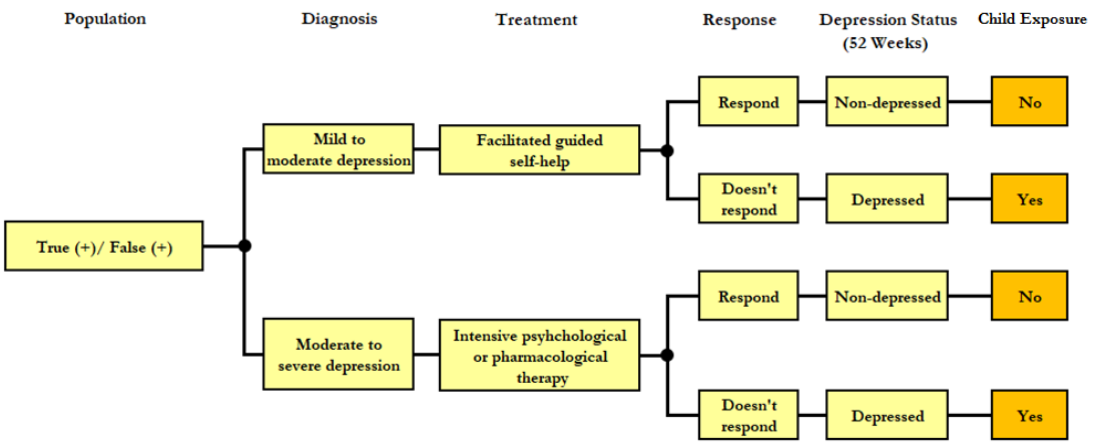


Figure 2.5: True positive and false positive treatment components in the decision analytic model used to assess the cost-effectiveness of screening for postnatal depression in chapter six.

Following the conclusions from this chapter, the specific objective for the empirical research example described over chapters three to six of this thesis is:

> To estimate the lifetime incremental effect of childhood exposure to symptoms of postnatal depression on the following decision endpoints: (i) QALYs (ii) costs incurred by the National Health Service & Personal Social Services, (iii) consumption costs and benefits on budgets outside of health and (iv) economic productivity.

Expanding on methodologies for indirect estimation, chapters three, four and five estimate lifetime evidence using observational data from two longitudinal birth cohort studies. Chapter six incorporates the evidence in an applied cost utility analysis.

Chapter 3: Postnatal Depression and Child Development

3.1 Summary

Within the overall context of the thesis, the purpose of the research described in this chapter is to carry out the first stage of the indirect estimation process by observing the impact of early life circumstances on intermediate child development outcomes. The research uses an applied empirical example exploring the relationship between maternal postnatal depression and child development. It firstly examines how the timing and chronicity of maternal depression symptoms are associated with socioemotional and cognitive development outcomes in children aged three to eleven and secondly specifies the additional lifetime evidence that is required when assessing the cost-effectiveness of the health intervention of post-natal screening. The final element of the chapter discusses appropriate methodologies for the first stage of indirect estimation which could be used to inform the design of future research. The following three methodological challenges are addressed: the different causal pathways through which early life circumstances can influence lifespan development outcomes; the need to observe long term intermediate outcomes at several observation points during childhood; and the selection of appropriate outcomes on which to measure lifespan development outcomes. Figure 3.1 summarises how the empirical evidence from this chapter contributes to the estimation of lifetime effects later in this thesis.

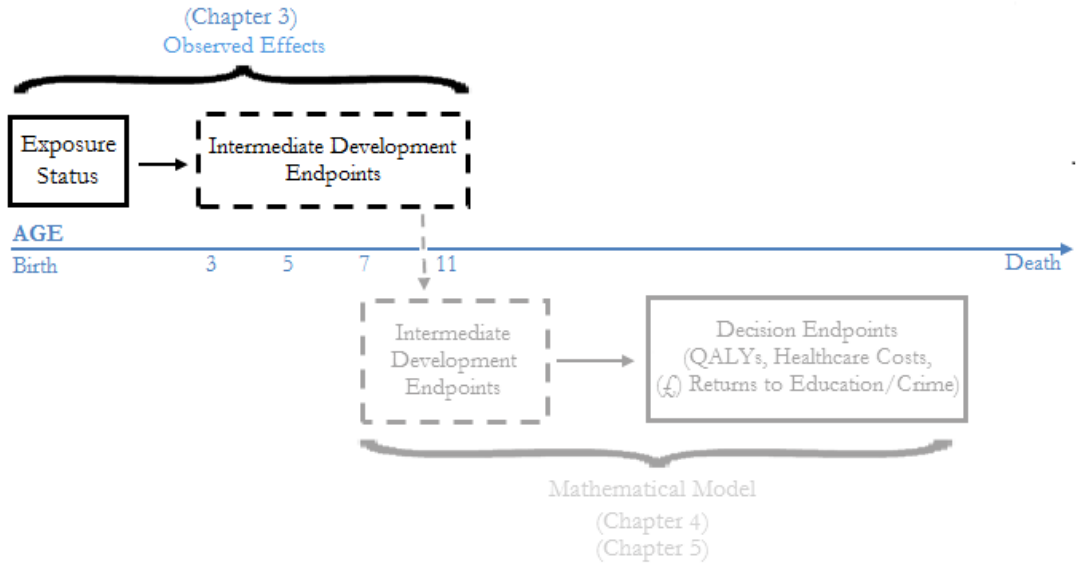


Figure 3.1: Demonstrates how evidence from this chapter contributes to the overall estimation of lifetime effects.

3.2 Introduction

3.2.1 Maternal Depression

Postnatal depression specifically diagnosed within the first year of childbirth, is highly prevalent and affects between 10-15% of new mothers (Grace et al., 2003). Symptoms are not distinct from major depressive disorder and include low mood, loss of energy, self-loathing, sleep disturbances, feelings of helplessness and despair, and in severe cases suicidal or infanticidal thoughts (Beck and Alford, 2009), (Goodman, 2004).

As described by Stewart et al. (2003), there are a variety of factors which may increase the risk of postnatal depression including: pre-existing mental health conditions; a genetic susceptibility for depression; pregnancy problems e.g. increased stress and hormonal changes leading to antenatal depression and anxiety; distress and trauma occurring as a result of obstetric complications; socioeconomic factors; social isolation; maternal attributes including reduced cognitive ability and a highly neurotic personality type; and unusually temperamental infants causing increased childcare stress and lack of sleep.

While most mothers enter remission twelve months following the onset of postnatal depression, symptoms can persist in up to 36-49% of cases (Vliegen et al., 2014). Therefore, postnatal depression represents a significant risk factor for future mental health conditions and may predict the onset of chronic depression (Vliegen et al., 2014). Following the postnatal period, the course of symptoms can vary in severity, timing, chronicity and duration given the heterogeneous nature of mental health disorders (Hammen and Brennan, 2003), (Vliegen et al., 2014).

The detrimental effects of maternal depression may extend beyond mothers to their children. Depression symptoms can affect interpersonal functioning (Kroenke et al., 2001) meaning that depressed mothers may be less likely to form secure attachment relationships with their infant which are important for child development (Broth et al., 2004), (Cornish et al., 2006), (Grace et al., 2003). Insecure attachment is linked to negative personality traits, poor peer relationships, reduced cognitive function, behavioural problems and mental health issues later in childhood (Weinfield et al., 2000), (Monique van Londen et al., 2007). The association between maternal depression and child development is well documented in theoretical and empirical literature (Cummings and Davies, 1994), (Grace et al., 2003), (Murray and Cooper, 1997).

3.2.2 Causal Pathway: Latency and Accumulation Models.

Two prominent hypotheses from the development literature aim to describe the causal pathway between maternal depression and child development. Firstly, the latency model indicates that developmental effects could occur if children are exposed to negative stimuli during a critical period, usually early in life, resulting in permanent biological changes (Power and Hertzman, 1997). An example of the latency model is provided through the effects of alcohol on the developing foetus, which increases the risk of genetic deformities if exposure occurs during a specific (critical) period in pregnancy (Rice and Barone Jr, 2000).

The latency model has been recognised in child attachment theory where it is suggested that a critical period of attachment occurs during the first two years of life (Bowlby, 1978). A large body of empirical evidence has assessed the effects of postnatal depression symptoms occurring during this supposed critical period. As indicated in systematic reviews by Kingston and Tough (2014) and Sanger et al. (2015), associations were identified between postnatal depression and measures of development in cognitive and socioemotional domains. However, these findings are not consistent across all studies.

The accumulation hypothesis is a second model that could explain the link between maternal depression and child development. This model assumes that the effects of negative exposures are cumulative (Power and Hertzman, 1997) (i.e. repeated exposures add up over time) which might be thought of similarly to the increased risk of cancer resulting from repeated cigarette smoking. The accumulation model may account for the effects of poor socioeconomic conditions on child development as children are repeatedly exposed to detrimental early environments (Power and Hertzman, 1997). Whilst only a small number of studies have investigated the accumulation hypothesis, findings consistently identify a significant negative association between chronic symptoms of maternal depression and both cognitive and socioemotional development (Brand and Brennan, 2009), (Sanger et al., 2015) (Bell, 2014).

3.2.3 Identifying Longitudinal Effects using Growth Curve Models

The effects of postnatal/maternal depression (as with the effects associated with early life circumstances in general) on lifespan development outcomes have the potential to change over time: early environments can have increasing, cascading or snowballing developmental effects if they influence fundamental abilities that are important determinants of future abilities (Heckman, 2008) (Masten and Cicchetti, 2010). On the other hand, decreasing effect sizes might occur if the detrimental effects of significantly adverse environments can be ameliorated by later beneficial factors such as strong family relationships, safe neighbourhoods and effective

schooling (Masten and Reed, 2002). This process is more commonly known as developmental "catch-up" and is exemplified by Rutter (1998) who found Romanian orphans with significant deficiencies in early physical and cognitive development achieved similar outcomes to British controls after being placed in nurturing UK foster homes.

The potential for lifespan development effect sizes to change over time has implications for research design. Typical studies collecting outcomes at one measurement interval may not be appropriate as they can provide no information regarding temporal changes to effects sizes after the study endpoint. Single outcome studies could lead to inappropriate assumptions if used to estimate lifetime effects. For instance, it is possible that a study which observes outcomes at a single observation might identify two early life circumstances with equivalent effects sizes on intermediate endpoints. Without any further information, these effect sizes would be assumed to extrapolate to equivalent effect sizes on lifetime endpoints. But this might not be appropriate if, say, one of the intermediate effects was increasing whilst the other was decreasing over time.

Temporal changes in effect sizes could ultimately place limitations on the reliability of policy recommendations from economic evaluation. For example, Hollingworth et al. (2012) evaluated the cost-effectiveness of different lifestyle intervention to treat childhood obesity by estimating the effectiveness of intervention on children's BMI, and extrapolating BMI trajectories to estimate lifetime healthcare costs. Hollingworth et al. (2012) suggest that the study may have either under or over exaggerated the cost-effectiveness of intervention as treatment effects had the potential to change over time and effectiveness evidence was obtained from short term RCTs (<2 years) with outcomes measured at a single follow up interval.

The indirect estimation of lifetime effects is likely to be more appropriately informed through studies designed to identify effects on repeated developmental outcomes in time series/longitudinal databases. Growth curve models are a group of analytical techniques applied to time series data which can identify changes in developmental effects over time (Curran et al., 2010). Specifically, growth curve models establish inter-individual (between people) differences in intra-individual (within person) change (Curran et al., 2010):

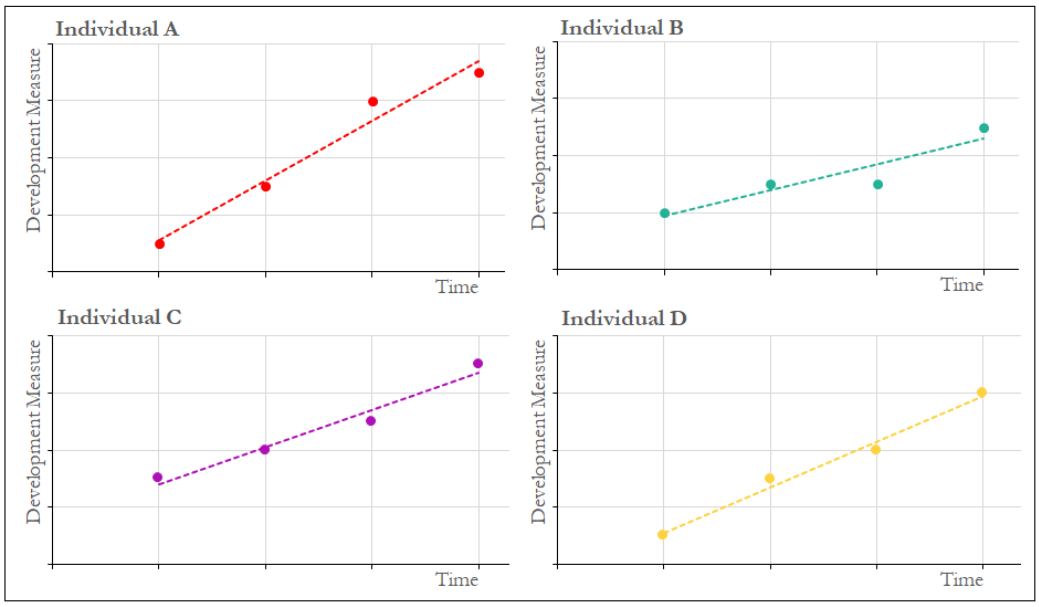


Figure 3.2: Displays development trajectories for four hypothetical children using four observations within each individual over time. For simplicity the line of best fit has only linear terms, but these can be extended to quadratic, cubic, exponential etc. depending upon the function of change.

Initially, growth curve models use longitudinal data to identify development trajectories for each child in a sample. Development trajectories are latent (unobserved) functions describing the trend in development outcomes *within each child* over time. A line of best fit is used to calculate each individual development trajectory as illustrated in Figure 3.2. Growth curve models then calculate an average development trajectory across all individuals within a sample by estimating the mean intercept and mean gradient of change, as shown in Figure 3.3. In growth curve modelling the mean trajectory is also known as the fixed effect.

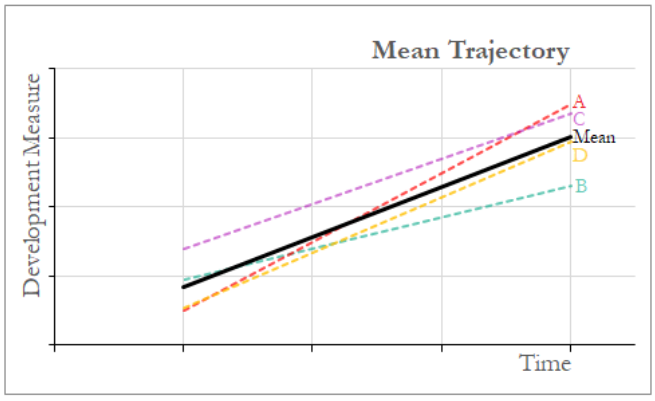


Figure 3.3: Illustrates the mean (fixed effects) development trajectory for four hypothetical children. Between individual changes relate to the variance (random effects) between individual and mean development trajectories in terms of intercept and/or gradient.

As with linear regression techniques, explanatory variables are introduced into growth curve models to explain the variability between individuals and the mean. Residual differences *between* individual trajectories and the mean development trajectory are known as random effects and these occur both in terms of variability in intercept and slope (gradient of change over time). For example, an explanatory variable such as gender might be introduced into a model to determine whether, on average, boys achieve better or worse initial development outcomes than girls (intercept), and whether their rate of change (gradient) is expected to be faster or slower over time, Figure 3.4.

By including explanatory variables capturing children's exposure status to symptoms of maternal depression, growth curve models can account for potentially changing effect sizes over time and provide an appropriate method to inform indirect estimation in this research. In addition to informing the thesis objectives, this form of analysis adds new evidence to the existing postnatal/maternal depression literature base. Whilst there are several longitudinal studies estimating the effects of postnatal depression on child development into adolescence there is less frequent evidence regarding the longitudinal effects of chronic maternal depression in studies with large sample sizes (n>1000) (Sanger et al., 2015).

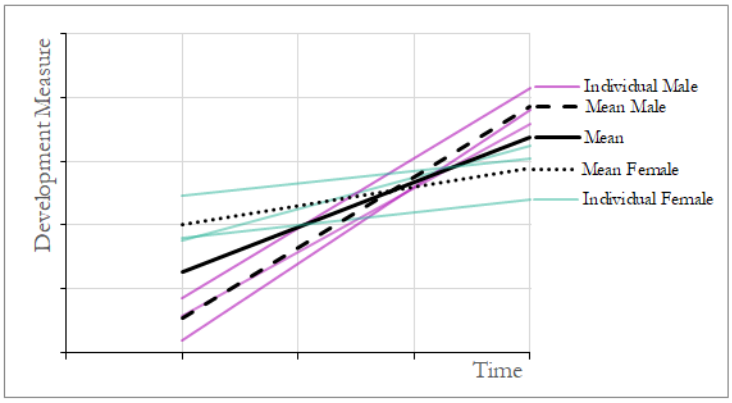


Figure 3.4: A hypothetical growth curve model that illustrates between person differences in within person change that are explained through gender. In this example males tend to achieve initially lower developmental outcomes but have a steeper gradient of change when compared with females.

3.2.4 Empirical Study Aims and Research Questions

The aim of this first stage of the indirect estimation process is to explore the relationship between symptoms of maternal postnatal depression on intermediate child development outcomes and investigate latency and accumulation hypotheses in a large longitudinal birth cohort.

This study determines whether the timing and chronicity of depressive symptoms differently affects the association between maternal depression and children's cognitive and socioemotional development by estimating growth curve models in time-series data across four observation points from infancy (age 3) to early adolescence (age 11).

The chapter also focuses on the impact of maternal depression symptoms later in childhood given their potentially confounding effects. Postnatal depression is a risk factor for future depression. Later episodes of maternal depression are not always accounted for in the current literature which may explain some of the inconsistent findings in the empirical evidence. This chapter attempts to isolate the effects of postnatal depression on child development such that evidence may be appropriately used to estimate lifetime effects through indirect estimation (chapter five). The chapter also provides recommendation for policy and identifies future research directions.

Given the aims above, the following research questions were devised:

1. Do symptoms of maternal depression affect children's cognitive and socioemotional development through a latency and/or accumulation model?
    - Are symptoms of *postnatal* depression negatively associated with cognitive and socioemotional development up to age 11?
    - Are cumulative symptoms of maternal depression *after* the postnatal period negatively associated with cognitive and socioemotional development up to age 11?
    - Does the size of association between symptoms of postnatal/maternal depression and children's cognitive and socioemotional development change over time?
2. Does the timing of symptoms have an impact on the size of association between maternal depression and cognitive and socioemotional development up to age 11?

3.3 Methods

3.3.1 Study Sample

Data was obtained from the Millennium Cohort Study (MCS), a longitudinal birth cohort study collecting information on 19,000 children born in the UK from 2000-2001 (UoL., 2017). The MCS has a disproportionately stratified sample to represent ethnic minorities and the four countries within the UK (Plewis et al., 2007). At the date of this research the MCS had administered five surveys when children were nine (months), three, five seven and eleven years

old. Each of the sweeps obtained detailed information on children's health, socioeconomic and family circumstances through parental questionnaires. Sweeps 2-5 also collected children's scores across several developmental measures within cognitive, socioemotional and physical domains. Parents and immediate family members were also questioned in each survey including items identifying their health, education, employment, living situation, parenting practices and well-being (Plewis et al., 2007).

In addition to the core MCS database, this analysis merged academic attainment data for Key Stage One and Key Stage Two (sweep five) from the MCS Linked Education Administrative Datasets (UoL, 2015a), (UoL, 2015b). Academic attainment data was available for 8,500 cohort members from England whose parents had consented to data linkage. The available variables include an average point score across all English National Curriculum SATs tests and individual point scores within each tested subject (Johnson and Setakis, 2015).

The study sample in this research was limited to natural mothers as children who are not biologically related to their parents may have complicated attachment relationships which might not be fully accounted for by variables available in the MCS: For example, adopted children are often at an increased risk of forming insecure attachment relationships when compared with non-adopted children (Van Den Dries et al., 2009). The study was also limited to analysing a maximum of one child from multiple births (i.e. twins/triplets) to avoid double counting effects for individual mothers.

3.3.2 Variables

3.3.2.1 Socio-emotional Development

Socioemotional development refers to children's capacity to form relationships with others and their ability to regulate and appropriately express emotions (Halle and Darling-Churchill, 2016). As suggested by Halle and Darling-Churchill (2016) socioemotional capacities might be characterised by four (overlapping) subdomains: *social competence* describes the ability of children to successfully interact with others; *emotional competence* is the ability to understand the emotions of others; *behaviour* relates to the ability of children to act in an appropriate way consistent with their emotions; and *self-regulation* relates to attention and control of behaviour.

It may not be appropriate to obtain a direct measure of socioemotional development from children as they are not good at objectively recognising and reporting their relationships and/or emotions (Atkins-Burnett and Meisels, 2001). Instead, proxy measures provide a useful method of assessment and are usually obtained from the observations of individuals close to the child e.g. parents, siblings and teachers (Atkins-Burnett and Meisels, 2001).

Observational measures of socioemotional development are often based on assessments of child behaviour: as outlined above, behaviour is an important socioemotional construct but may also reveal characteristics within other subdomains as social, emotional and self-regulatory competencies can manifest in chosen behaviours (Atkins-Burnett and Meisels, 2001). Behavioural measures are sometimes dichotomised as either externalising behaviours such as tantrums and aggressiveness which might indicate adaptive emotional problems, or internalising behaviours including social withdrawal, worrying, and excessive clinging which may be indicative of affective disorders such as depression or anxiety (Atkins-Burnett and Meisels, 2001), (Halle and Darling-Churchill, 2016).

Here, socioemotional development was measured using maternal responses to the Strengths and Difficulties Questionnaire (SDQ). The SDQ is a behavioural screening questionnaire for children aged 3-16, consisting of twenty-five items each scored from 0-2 (Flouri et al., 2014), (Goodman, 2001). While the SDQ was primarily designed for the identification of childhood mental health conditions (Vliegen et al., 2014), it has been applied in multiple peer reviewed studies as the primary assessment for socioemotional development e.g. Hintermair (2006), Kelly et al. (2012), McMunn et al. (2011), and Rogers et al. (2012).

The SDQ has multiple scales which include: the externalising subscale with ten items relating to behaviours such as aggression, impulsiveness and lack of control; the internalising subscale containing ten items assessing control over thoughts where problems may manifest as social withdrawal, depression and anxiety; and the total difficulties scale a continuous measure scored from 0-40 on twenty items, these being the sum of the externalising and internalising subscales (Goodman, 2001).

As this empirical research is primarily concerned with the effect of maternal depression on global child development this study used the SDQ total difficulties score as the primary outcome measure providing the widest assessment of socioemotional behaviours. Secondary results were obtained for both the externalising and internalising subscales to establish whether effects differ according to different types of behaviour. All socio-emotional outcomes were measured at four distinct time points for MCS sweeps 2-5.

3.3.2.2 Cognitive Development

The cognitive domain relates to the development of children's mental capacities. Cognition is defined by Bjorklund and Causey (2017) as the process by which knowledge is acquired and manipulated. As cognition is a mental process it cannot be directly measured but can be inferred by assessing an individual's performance on tasks requiring cognitive function (Bjorklund and Causey, 2017). Global cognition is difficult to assess as it includes a wide range

of diverse capacities (Bjorklund and Causey, 2017) and it is not immediately evident which are most significant. For example, researchers might be equally interested in the development of specific cognitive abilities such as auditory perception, decision making, intelligence, learning, memory, processing speed, quantitative knowledge, reading, retrieval, visual perception and/or writing (Alfonso et al., 2005).

The field of psychometrics is directed at understanding and providing appropriate measures of cognition. A pioneering theory in this field was Spearman's g-factor hypothesis which suggested an overarching cognitive ability exists that subsumes all other narrower abilities (Jensen, 1998). This theory explained empirical evidence that identified high correlations between tests assessing narrower cognitive abilities. It also offered an explanation as to why bright individuals tend to repeatedly outperform less bight individuals across a wide range of academic testing scenarios (Jensen, 1998).

The idea of an overarching *general cognitive ability* has been adopted in modern cognitive theories including Carroll's three stratum theory and Cattell-Horn-Carroll (CHC) theory (Alfonso et al., 2005). General cognitive ability (GCA) forms the basis of many psychometric tests and measures of cognitive development. One example is the British Ability Scales II which test children on different scales for verbal, non-verbal reasoning and spatial abilities. A group of statistical techniques termed factor analyses are used by psychometric tests to derive an overall measure of GCA from scores on underlying subscales (Hill, 2005), Figure 3.5.

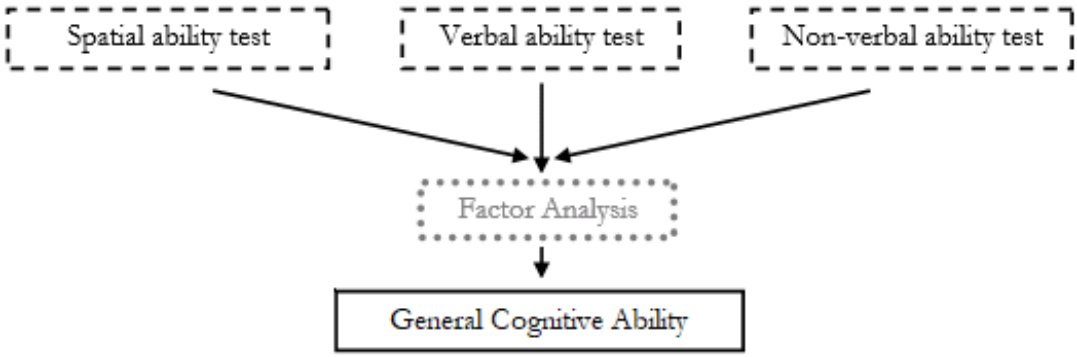


Figure 3.5: Illustrates the derivation of general cognitive ability from tests of narrower cognitive abilities using factor analysis. Image adapted from Hill (2005)

Table 3.1: Summary of Variables used to Identify General Cognitive Ability

Measure	Description
Sweep 2	
BAS Naming Vocabulary[1]	A verbal scale which assesses spoken vocabulary. Children are required to identify coloured objects from a booklet.
Bracken School Readiness Assessment[1]	Evaluates children's readiness for primary school by assessing their understanding of 88 basic concepts.
Sweep 3	
BAS Naming Vocabulary[1]	As sweep 2.
BAS Picture Similarities[1]	Measures problem solving ability. Children are given a picture and asked to place it below the most similar picture from a row of four.
BAS Pattern Construction[1]	Tests spatial awareness by assessing children's ability to fit together flat squares or solid cubes with different patterns on them.
Sweep 4	
BAS Pattern Construction[1]	As sweep 3.
BAS Word Reading Score[1]	Assesses reading ability. Children are asked to read a series of increasingly difficult words aloud.
National Curriculum Key Stage One[2]	The average of points scored in academic tests for Key Stage 1 across Reading, Writing, Maths and Science.
NFER Progress in Maths Test [1]	A shortened version of the National Foundation for Education Research (NFER) standard Progress in Maths test. Asks children to complete mathematical calculations using pencil and paper.
Sweep 5	
BAS Verbal Similarities[1]	Measures verbal abilities. Interviewers read three words to the child and the child must say how the words are related.
CANTAB Spatial Memory Task[3]	Tests children's ability to retain information and manipulate remembered items in their working memory. Children are asked to find a blue token by process of elimination using a computer.
National Curriculum Key Stage Two[2]	The average of points scored in academic tests for Key Stage 2 across English, Maths and Science.

Notes: Full details and descriptions are provided in; Connelly (2013)[1]; Johnson and Setakis (2015)[2]; and Atkinson (2015)[3].

Several psychometric testing subscales were administered to participants in the MCS which varied in number and type between sweeps. As no psychometric tests were administered in full it was not possible to derive GCA using a validated procedure. Connelly (2013) & Jones and Schoon (2008) advocate the use of Principal Component Analysis (PCA), a form of factor analysis, to identify children's GCA using the measures that are available in the MCS. The process of PCA uses orthogonal transformation (a form of transformation applied in linear algebra to matrices) to convert a set of correlated variables to a smaller set of uncorrelated principle components (Shlens, 2014). PCA might be considered appropriate for the cognitive

variables in the MCS as each assess a higher-level function that is likely to contribute to GCA and are commonly included constructs in validated psychometric tests (Hill, 2005).

This study extended the methods suggested by Connelly (2013) and Jones and Schoon (2008) by incorporating academic attainment scores for Key Stage 1 and Key Stage 2 as variables informing PCA. The inclusion of academic attainment variables was considered appropriate as GCA is strongly predictive of performance on school as well as psychometric tests (Jensen, 1998). The academic attainment variables included in this analysis were mean points scores across subjects of English, Maths and Science – average performance across these subjects is thought likely to assess a range of higher-level cognitive abilities like those assessed in psychometric tests. In total the analysis conducted PCA at four separate occasions to identify four different GCA variables for sweeps 2-5. Details of all the variables used to identify GCA are provided in Table 3.1.

3.3.2.3 Maternal Variables

Symptoms of maternal depression/psychological distress were identified at each of the five sweeps. For sweep one, the MCS collected mothers' responses to the reduced 9-Item Rutter Malaise Inventory. This asked mothers if, in general, they felt tired, miserable, worried, violent, scared, irritated, jittery, or nervous and whether they often suffer from a racing heart. Positive responses to each item were given a score of one, and a total score was derived by summing the responses. A threshold of four or more was used to identify *symptoms* of postnatal distress (PND) (Dex and Joshi, 2004).

Symptoms of maternal distress were identified in sweeps 2-5 based on questionnaire responses to the Kessler-6 index which asked mothers how often they feel hopeless, restless, that everything is an effort, worthless and nervous. Each item is scored from zero "none of the time" to four "all of the time" according to the frequency at which the item occurred. A total score ranging from 0-24 was obtained by summing each item. A threshold of thirteen or above was adopted to indicate the presence of significant psychological symptoms (Prochaska et al., 2012). Both the 9-Item Rutter Malaise Inventory and the Kessler-6 Index are detailed in Appendix 3.1

3.3.2.4 Child and Family Variables

The MCS provided a variety of child and family characteristics that were used as covariates in the statistical models. As this was a longitudinal study, the covariates could either be defined as time invariant or time variant. Time invariant covariates were assumed to remain constant over the study period (i.e. their values could not change for subsequent sweeps). These

were obtained during the birth sweep and included the biological sex of the child, the child's ethnicity and the total number of siblings the child had at birth. In contrast, time variant covariates are variables that could change across sweeps and were therefore collected at four observation points between sweeps 2-5. These included the child's age (in months) at SDQ assessment, the mother's employment and education status, and low family income status. All covariates were derived from the MCS parental response questionnaires and are detailed in Table 3.2. The choice of covariates was informed though a literature review conducted by Bell (2014), who identified variables in the MCS related to the individual and their environment that were likely to influence cognitive and socioemotional development. The inclusion of these variables is consistent with the theoretical literature reviewed in chapter two.

Table 3.2: Summary of Child and Family Covariates

Measure	Description
Time Invariant (Sweep 1)	
Biological Sex	Dichotomous variable identifying the child's biological sex at birth (male/female).
Ethnicity	Categorical variable grouping a child's ethnicity at birth as either: White British; Mixed Race; Indian; Pakistani or Bangladeshi; Black; or Other.
Siblings	Categorical variable indicating the total number of siblings a child had at birth. The four categories were: zero; one; two; three or more.
Maternal Age at Birth	Continuous variable identifying the age of the mother at child birth.
Time Variant (Sweeps 2-5)	
Age	Continuous variable identifying the age of a child (to the nearest month) at each SDQ assessment.
Mother's Employment	Dichotomous variable identifying whether the mother is in employment (yes/no).
Mother's Education	Categorical variable for the mother's highest education level achieved at each sweep. Education is grouped by UK National Vocational Qualification Scale: None; NVQ level 1; NVQ level 2; NVQ level 3; NVQ level 4; NVQ level 5; overseas qualification
Family Income	Dichotomous variable identifying whether the family income falls below 60% of the UK median (yes/no).
Single Mother	Dichotomous variable identifying whether the mother is in a relationship with another individual living in the child's household (yes/no).

3.3.2.5 Omitted Variables

Several other variables could not be included as covariates in the growth curve models as they were not measured in the MCS dataset, comprising: maternal/child genetic markers, maternal depression/mental health symptoms during the antenatal period, and maternal cognitive/personality assessments. Each of these variables may potentially confound the

observed relationship between PND and child development as they are identified risk factors for maternal PND (Stewart et al., 2003) which could separately predict effects on child development outcomes.

An additional two potentially confounding variables were measured in the MCS but were not included as covariates to avoid over specifying/complicating the analysis. Infant temperament was measured on the Carey Infant Temperament Scale at the 9-month sweep, and whilst likely to strongly predict children's later development was considered unlikely to have a strong confounding effect given the classification of infant temperament as a relatively minor risk factor for PND (Stewart et al., 2003). Paternal depression symptoms (measured equivalently to maternal symptoms throughout the MCS) are likely to be correlated with maternal depression symptoms (Nutall et al., 2019) and are detrimentally associated with children's later development (Ramchandani et al., 2005). However, Nuttall et al. (2019) suggest the interdependent nature of maternal and paternal PND requires dyadic data analytic models if including both as covariates in a single regression equation – the application of dyadic data models within an already complex growth curve modelling specification was beyond the scope of this analysis.

3.3.3 Statistical Methods

3.3.3.1 Underlying Growth Curve Models

Growth curve models were fit to the data using a multi-level linear regression framework: multi-level regression models are designed to account for different levels of hierarchical data where observations are nested within groups (Curran et al., 2010). Typical linear regression models do not account for nested data as they only incorporate one source of residual variance. In contrast, multiple residuals are included in the multilevel framework at each hierarchical level (Steele, 2008). For instance, the multi-level framework was originally developed to analyse educational data, assuming that children are nested in classrooms which are nested in schools, by including three sets of residuals one each for children, classrooms and schools (Curran et al., 2010).

Growth curve models can be fit as multi-level regression models by assuming time series data is hierarchical. This requires each repeated outcome to be nested within individual children, thus incorporating residuals for both repeated outcomes and for children (Curran et al., 2010). This research conducted growth curve analysis using multi-level regression for GCA and SDQ outcomes separately. Both models included four observation points at sweeps 2-5.

A square root transformation was applied to the SDQ outcome variable given its positive skew. Justification for this transformation is based on the argument that skewed outcome variables can lead to imprecise or biased regression coefficients (Jones, 2010), further justification is described in the results section when conducting residual diagnostics. All analyses were conducted using the computer software STATA.

The growth curve models were fit in several stages as recommended by Curran et al. (2010). First, an underlying fixed effects growth function (mean development trajectory) was fit by averaging effects across individual development trajectories. The estimation of a fixed effects growth function required specification of a time metric (i.e. a variable to denote when an observation occurred). For SDQ outcomes the time metric was defined as the child's age when outcomes were measured. Likelihood ratio tests were used to determine whether the fixed effects function was best represented by constant, linear, or quadratic terms for age. For GCA, no single time metric could be defined as multiple tests were administered at different intervals. Therefore, time was entered as a categorical variable using the average age of all children within each of the sweeps.

Second, random effects terms for the intercept and for the gradient were fit which allowed the covariance between an individual's observations to depend on the timing of the measurement (Steele, 2008). Likelihood ratio tests were used to assess the significance of each random effect. At this stage each model was also tested using unstructured, autoregressive and exponential covariance structures. The most appropriate covariance structure was selected as the one with the lowest Akaike Information Criteria (Steele, 2008).

Next, between-individual differences were identified by adding child and family covariates to the underlying growth function, thus establishing a conditional growth curve. It was anticipated that differing covariate categories might non-identically influence the rate of future development. For example, whilst the average development trajectory for boys and girls is likely to follow a similar shape over time, it is unlikely that the gradient of change is identical. To allow for these differences, interaction terms were tested between each covariate and the time metrics in the model. Likelihood ratio tests were used to identify significant interactions and non-significant interactions were dropped from the model on the grounds of parsimony.

Finally, two separate growth curve models were estimated for each of the GCA and SDQ outcomes which differed in terms of the specification of the maternal distress variable (described below). Two separate specifications were designed as this allowed investigation of different research questions relating to (i) the pathway of effect between maternal depression and child development (latency or accumulation hypotheses) and (ii) whether the timing of exposure was important. Interactions were assessed between each maternal distress variable and

the time metrics in both model specifications. Further detail on growth curve modelling using multilevel regression is described in Curran et al. (2010) and Steele et al. (2008).

3.3.3.2 Specification of Growth Curve Model 1

Growth Curve Model 1 addresses the first research question – to determine whether symptoms of maternal depression affected children's cognitive and socioemotional development through a latency and/or accumulation model.

The latency model was investigated by including a time invariant explanatory variable to determine the influence of postnatal depression symptoms on child development outcomes. Symptoms of postnatal distress (PND) were assigned if mothers scored above threshold on the 9-Item Malaise Inventory during sweep one.

A second explanatory variable evaluated the accumulation hypothesis and was thus named *cumulative distress*. This time-variant variable was obtained by summing the total number of times that mothers had scored above threshold on the Kessler 6 index between sweeps 2-5. Each child's outcome was adjusted for cumulative distress up to the time of the observation but not after e.g. outcomes at sweep four were adjusted for accumulated symptoms of distress for sweeps two, three and four, but not adjusted for symptoms at sweep five. For more clarity Table 3.3 illustrates (hypothetical) maternal distress trajectories and the associated value for the cumulative distress variable at each sweep.

Table 3.3: Example Coding Scheme for the Cumulative Distress Variable

Maternal Depression Trajectory (Sweeps 2-5)	Value of Cumulative Distress variable			
	Sweep 2	Sweep 3	Sweep 4	Sweep 5
$N_2N_3N_4N_5$	0	0	0	0
$D_2D_3D_4D_5$	1	2	3	4
$D_2N_3D_4N_5$	1	1	2	2
$N_2D_3N_4D_5$	0	1	1	2
$N_2N_3D_4D_5$	0	0	1	2

Key: N_2=No depression symptoms during sweep 2
D_3=Depression symptoms during sweep 3

Notes: Cumulative Distress variable is time variant and can take different values at each sweep. The value of the variable is obtained by summing total instances of depression symptoms up to the current sweep, beginning at sweep two: For example, a mother with depression trajectory $D_2D_3D_4D_5$ would have cumulative distress equal to 1 at Sweep 2 (D_2); 2 at Sweep 3 ($D_2 + D_3$), 3 at Sweep 4 ($D_2 + D_3 + D_4$) and 4 at Sweep 5 ($D_2 + D_3 + D_4 + D_5$).

The β coefficient attached to the cumulative distress variable was interpreted as a continuous regression coefficient: it identified the impact per each additional episode of maternal distress. Therefore, a significant β coefficient would predict that more frequent episodes of distress were

associated with larger effect sizes, providing evidence of a dose dependent relationship between symptoms of distress and SDQ/GCA outcomes. The coding for the cumulative distress variable follows methods by McCoach and Kaniskan (2010) who use growth curve models to evaluate the impact of an intervention on longitudinal measures of children's reading fluency.

3.3.3.3 Specification of Growth Curve Model 2

Growth Curve Model 2 estimates the association between children's GCA/SDQ scores and different trajectories of maternal distress, a method commonly used in the literature. Allocating mothers to trajectories is a method commonly used in the literature (for examples see Brennan et al. (2000) and Vliegen et al. (2014)) and enabled comparison of different exposure groups for combined postnatal and maternal distress symptoms. This was useful for objective two, which assessed whether the timing of depression symptoms was important.

Trajectories were initially defined by splitting mothers into groups according to whether they had symptoms of PND or not. It was then assumed that mothers followed trajectories of no, episodic, or persistent symptoms between sweeps 2-5. Episodic trajectories were defined (arbitrarily) as infrequent symptoms and were assigned if mothers scored above threshold one or two times on the Kessler 6 Index between sweeps 2-5. Persistent trajectories were frequent symptoms assigned if mothers scored above threshold three or four times between sweeps 2-5.

The six maternal depression trajectories are described below:

1. *No symptoms* if mothers were below threshold across all sweeps
2. *PND symptoms & no further symptoms* if mothers were above threshold during sweep 1, and below threshold for sweeps 2-5
3. *No PND symptoms & episodic depression* if mothers were below threshold during sweep 1, and above threshold either one or two more times for sweeps 2-5
4. *PND symptoms & episodic depression* if mothers were above threshold during sweep 1, and above threshold either one or two more times for sweeps 2-5
5. *No PND symptoms & persistent depression* if mothers were below threshold during sweep 1, and above threshold either three or four more times between sweeps 2-5
6. *PND symptoms & persistent depression* if mothers were above threshold during sweep 1, and above threshold either three or four more times for sweeps 2-5

Each maternal distress trajectory was specified as a time-invariant dummy variable which assumed that mothers followed the same trajectories throughout the study (i.e. mothers could not move from having persistent depression to episodic depression). Pairwise comparisons were used to identify significant effects between groups. The cumulative distress variable was not

included in Growth Curve Model 2 as later symptoms of distress were accounted for within the trajectories.

3.3.3.4 Interpreting Results

Parameter estimates, standard deviations and confidence intervals were obtained for each model with outcomes predicted for children aged three, five, seven, and eleven. These were estimated by appropriately centering the time metrics used in the model. For example, to predict outcomes for children aged three, the time metrics were coded as time equal to the age at SDQ outcome minus three (years). The methods used for making time-specific predictions in growth curve models followed recommendations by Biesanz et al. (2004).

As SDQ model predictions were made on the square root scale, where appropriate, incremental effects were back transformed onto the raw scale: average marginal effects were estimated in STATA using the "margins" command. These effects were then squared, to obtain the incremental differences on the raw scale.[3] Back transformed outcomes were compared to models where no transformation was applied to confirm their legitimacy.

3.3.4 Data Imputation

As data was obtained from a longitudinal study which predominantly used large self-report questionnaires it was anticipated that there would be a large amount of missing data. As missing data can result in parameter estimates being inefficient and biased (Briggs et al., 2003) this analysis used imputation methods to obtain complete databases. Multiple imputation by chained equations (MICE) was applied assuming that the data was Missing at Random (MAR) (Briggs et al., 2003), (White et al., 2011).

The MAR assumption requires the missing data to either occur randomly or for any systematic changes to be explained through characteristics within the observed dataset (Briggs et al., 2003). Whilst it is not possible to formally test the MAR assumption, it is likely that the non-random missing data in the MCS could be explained through a variety of the other child/family variables included in the analysis. For example, missing data might occur more often for lower income

[3] In some cases, back transformations may be biased if they do not include the addition of a constant known as a smearing estimator. Smearing estimators might be required if the back transformations affect the predicted mean, and these typically occur for multiplicative functions (e.g. exponential) or through additive effects of polynomial functions (e.g. square, cube etc.). Smearing estimators are not required for back transformations of incremental effects in square models as was applied in this analysis. Further information on smearing estimators is provided by Buntin et al. (2004) and Jones (2010).

families and this lower probability might be identified through other observed socioeconomic variables e.g. employment and education status.

The MICE procedure predicts missing values based on an individual's observed values using multivariate regression. Here, separate regression equations were estimated for all variables including outcomes. Where possible each equation incorporated observations for the missing variable from the previous and subsequent sweeps, for example, the regression equation to predict the SDQ outcome at sweep three included SDQ outcomes at sweeps two and four in the explanatory variable vector.

Predictive mean matching (PMM) was used on all SDQ scores and the 9- Item Rutter Malaise Inventory to ensure feasible values were imputed (White et al., 2011). The variables for cumulative maternal distress were derived after imputation to ensure that the value of each variable was possible according to the observed scores identified in previous sweeps (i.e. the imputed cumulative value could not decrease across subsequent sweeps).

The imputation process was repeated 10 times to generate 10 imputed databases. Each model was estimated on all of the imputed databases and combined into a single estimate using Rubin's law (Royston, 2004), (Rubin, 2004), (White et al., 2011). As suggested by Young and Johnson (2015), data was imputed for cases where observations were missing within sweeps and was also imputed for observations where individuals did not reply to entire sweeps.

Growth curve models do not require all observations within an individual to be present for that individual to be included in the analysis (Kwok et al., 2008), therefore the impact of missing data was likely to be minimal in this analysis. It was considered most appropriate to report primary results for the complete case database, and secondary results for the imputed databases: complete case analysis is often appropriate as the primary analysis where the impact of missing data is negligible (Jakobsen et al., 2017).

3.4 Results

3.4.1 Descriptive Summary

3.4.1.1 Response Rates

Responses were obtained from 14,501 mother-child dyads during sweep one of the MCS. Of these, 9,206 (65.46%) responded to all subsequent sweeps. The response rates between sweeps varied and was lowest for sweep five. A lower number of responses were recorded for the GCA outcomes as data was only available for English respondents, compared with UK response for the SDQ variables. Table 3.4 summarizes response rates for the non-imputed database across all sweeps in the MCS.

Table 3.4: Total Number of Questionnaire Responses

	Any Item	SDQ Response	GCA Response
Sweep 1	14,501	N/A	N/A
Sweep 2	13,305	12,957	5,619
Sweep 3	13,629	13,450	6,555
Sweep 4	12,564	12,404	6,803
Sweep 5	11,526	11,279	5,850
All Sweeps	9,206	8,830	4,618

3.4.1.2 Identifying General Cognitive Ability

There was a high level of correlation observed between psychometric and academic variables. The correlations within each sweep ranged from -0.20 (between the CANTAB Spatial Working Memory Tasks and BAS Verbal Abilities subscale at sweep five) to 0.72 (between BAS Word Reading Score and Key Stage 1 Average Points Score at sweep four). All within sweep correlation coefficients for the cognitive variables are reported in Table 3.5

Principal component analysis was performed separately on four occasions for sweeps 2-5. The PCA procedure condenses a vector of variables to a smaller vector of variables by analysing the correlations matrix between variables using linear algebraic matrices rotations (Steger et al., 2006). Variables making up the vector of condensed factors are termed *principal components*. Each PCA analysis across a vector of *n* variables produces *n-1* principal components and each principal component explains a different proportion of the total correlation between the original variables. It is typical for analysts to discard principal components that explain a small proportion of the total variance. Significant principal components are often identified if their associated Eigenvalue is greater than one, or by analysing a Scree plot of Eigenvalues (Steger et al., 2006).

Table 3.5: Within Sweep Correlation Matrices for all Cognitive Variables

Sweep 2

	Bracken SRA	BAS NV
Bracken School Readiness Assessment	1.00	
BAS Naming Vocabulary (NV)	0.56	1.00

Sweep 3

	BAS NV	BAS PS	BAS PC
BAS Naming Vocabulary	1.00		
BAS Picture Similarities (PS)	0.31	1.00	
BAS Pattern Construction (PC	0.36	0.36	1.00

Sweep 4

	KS1 AP	NFER Maths	BAS PC	BAS WR
Key Stage 1 Average Points (KS1 AP)	1.00			
NFER Progress in Maths	0.57	1.00		
BAS Pattern Construction (PC)	0.45	0.49	1.00	
BAS Word Reading (WR)	0.72	0.54	0.33	1.00

Sweep 5

	KS2 AP	BAS VS	CANTAB SP
Key Stage 2 Average Points (KS2 AP)	1.000		
BAS Verbal Similarities (VS)	0.45	1.00	
CANTAB Spatial Memory (SP	-0.41	-0.20	1.00

Table 3.6: Principal Component Analysis Results

	Eigenvalue	% of Variance Explained	Cumulative % of variance explained	Is solution a significant component?
Sweep 2 (n=12,990)				
Component 1	1.58	79	79	Yes
Component 2	0.42	21	100	No
Sweep 3 (n=14,277)				
Component 1	1.64	55	55	Yes
Component 2	0.72	24	79	No
Component 3	0.64	21	100	No
Sweep 4 (n=6,844)				
Component 1	2.57	64	64	Yes
Component 2	0.72	18	82	No
Component 3	0.45	11	93	No
Component 4	0.26	7	100	No
Sweep 5 (n=5,879)				
Component 1	1.72	57	57	Yes
Component 2	0.80	27	84	No
Component 3	0.48	16	100	No

Eigenvalues (Table 3.6) and scree plots (Appendix 3.2) in this analysis identified a single (unrotated) significant principal component for each of the sweeps two, three, four and five. This single component explained at least 50% of the variance between the cognitive variables within each sweep (Table 3.6).

It was assumed that each significant factor identified in the PCA represented GCA at each sweep. This assumption was justified as mean group GCA scores followed trends that are typical for cognitive variables in UK data (Jones and Schoon, 2008), where higher scores were associated with girls, children with higher family income, and children who were not from an ethnic minority group, Table 3.7. The face validity of GCA variables were further illustrated by the high level of correlation between variables for each of the sweeps, Table 3.8.

Table 3.7: **Between Sweep Correlation Matrix GCA**

	GCA Sweep 2	GCA Sweep 3	GCA Sweep 4	GCA Sweep 5
GCA Sweep 2	1.000			
GCA Sweep 3	0.5158	1.000		
GCA Sweep 4	0.4874	0.5995	1.000	
GCA Sweep 5	0.4466	0.5284	0.7273	1.000

The PCA results were consistent with those obtained by Jones and Schoon (2008) who also identified a single factor explaining correlations between cognitive variables in the MCS. This provided further vindication for their use as cognitive outcome variables here. All GCA variables were transformed to a common scale with mean=100 and standard deviation=15, which is often used for cognitive variables e.g. IQ.

Table 3.8: **Mean GCA Scores by Covariate Category**

	GCA Sweep 2	GCA Sweep 3	GCA Sweep 4	GCA Sweep 5
Biological Sex				
Male	98.06	98.87	99.36	99.93
Female	101.93	101.16	100.6	100.06
Family Income				
<60% median UK income	93.02	94.19	93.44	93.39
>=60% median UK income	103.35	102.99	103.017	102.75
Ethnicity				
White	101.64	101.39	100.82	100.71
Not White	89.22	92.57	96.84	97.05

3.4.1.3 Imputation Results

Missing values occurred through non-response to complete sweeps and non-response to items within sweeps. The number of participants with at least one missing value increased with each subsequent sweep and totalled 8,246 by sweep five. Multi-response variables GCA, SDQ, the 9-Item Rutter Malaise Inventory, and the Kessler-6 inventory were responsible for the largest amounts of missing data. Table 3.9 summarises the missing data in each variable across all sweeps.

Table 3.9: Total Number of Imputations per Variable per Sweep

Variable	Sweep 1	Sweep 2	Sweep 3	Sweep 4	Sweep 5
GCA	N/A	1392	551	333	1444
SDQ	N/A	1544	1051	2097	3222
Malaise Inventory Total Score	578	N/A	N/A	N/A	N/A
Kessler 6 Inventory	N/A	3393	2649	3765	5090
Family Income	N/A	1777	1996	3261	3793
Mother Employment Status	N/A	1758	1955	3254	4196
Single Mother	N/A	1787	1955	3254	3793
Maternal Age	3	N/A	N/A	N/A	N/A
Mother Education				No missing values	
Child Ethnicity				No missing values	
Child Gender				No missing values	
Child Age				No missing values	
Number of Siblings				No missing values	
At least 1 Imputation (all variables)	582	3661	5253	6611	8246

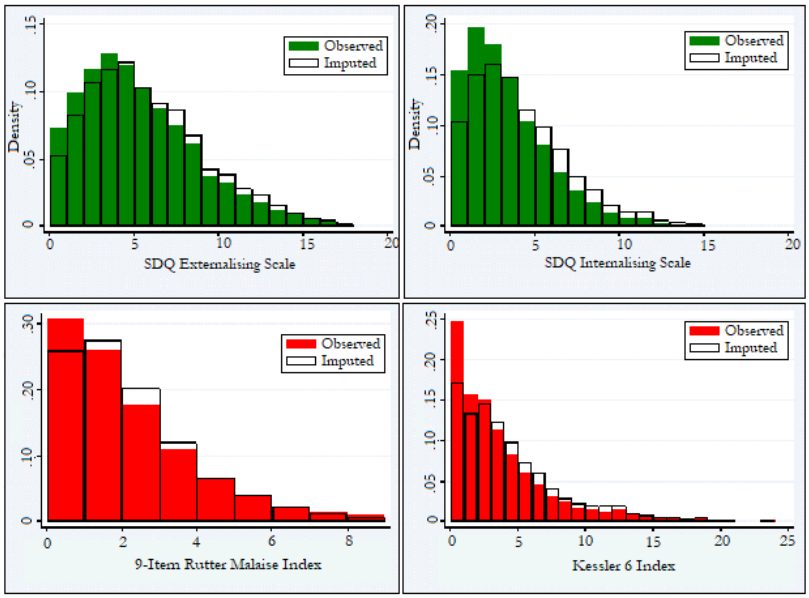


Figure 3.6*:* Histograms illustrating the distribution of observed (coloured) and imputed (clear) values for: sweep two SDQ externalising scale (top left); sweep two SDQ internalising scale (top right); sweep one Rutter 9-item malaise index (bottom left); and sweep two for the Kessler 6 index (bottom right).

3.4.1.4 Distribution of Covariates by Exposure Groups

When compared with mothers with no symptoms, children exposed to symptoms of PND were more likely to be from an ethnic minority group; had a higher proportion of mothers who were unemployed, not married and had no educational qualifications; and had larger families with lower household incomes. This combination of factors suggests that the exposure groups were imbalanced for important characteristics that may affect development, despite differences between individual characteristics being relatively small. Table 3.10 summarises the distribution of covariates for sweep one (if time invariant) and sweep two (if time variant) by exposure status to symptoms of PND.

Table 3.10*:* Baseline Characteristics by Postnatal Depression Exposure Status

	Child Not Exposed to PND Symptoms	Child Exposed to PND Symptoms
Total Number (n)	12,029	2,034
Child Biological Sex (%)		
Female	49.17	47.20
Male	50.83	52.80
Child Ethnicity (%)		
White	87.38	82.35
Mixed race	2.69	3.10
Indian	1.97	3.15
Pakistani or Bangladeshi	4.17	7.52
Black	2.84	3.00
Other	0.95	0.88
Total Number of Siblings (%)		
0	43.26	37.36
1	34.83	35.35
2	14.88	18.14
3+	7.02	9.14
Household Income[1] (%)		
<60% of UK Median	28.35	43.62
>=60% of UK Median	71.65	56.38
Mothers Education Qualifications[1] (%)		
NVQ 1	7.60	10.18
NVQ 2	28.91	29.96
NVQ 3	15.34	15.69
NVQ 4	31.04	20.46
NVQ 5	4.19	2.66
Overseas	2.12	3.74
None	10.79	17.31
Mother is Single Parent[1] (%)		
Yes	14.84	23.68
No	85.16	76.32
Mother is in Employment[1] (%)		
Yes	56.38	42.65
No	43.62	57.35
Mother's Age at Child Birth (mean)	28.77	27.59

Notes: Values for time variant covariates[1] were obtained at sweep two.

3.4.1.5 Trajectories of Mean Outcomes over Time

In general, SDQ total difficulties scores appeared to follow a curvilinear trajectory with respect to time decreasing between the ages of three and five, remaining relatively constant thereafter. As expected GCA scores remained relatively stable as variables for each sweep were centred with mean equal to 100. Children exposed to symptoms of PND had higher (worse) unadjusted mean SDQ total difficulties scores and lower (worse) unadjusted mean GCA scores than children whose mothers did not have symptoms of PND where incremental differences were equal to 3.20 (SDQ) & -3.78 (GCA) (sweep two), 3.08 & -2.31 (sweep three), 3.14 & -3.19 (sweep four), and 3.06 & -2.88 (sweep 5). Similarly, differences in unadjusted mean SDQ total difficulties and GCA scores were identified for children who were exposed to at least one episode of maternal depression during sweeps 2-5, Figure 3.7.

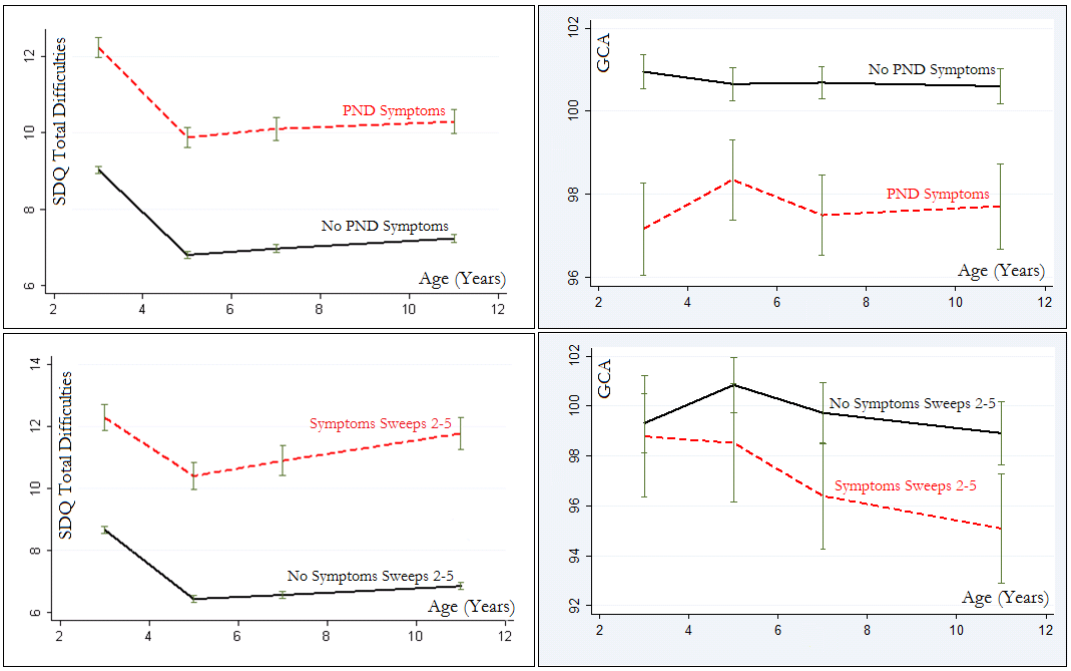


Figure 3.7: Unadjusted mean SDQ total difficulties (left) and GCA (right) scores for sweeps 2-5 and associated 95% confidence intervals by: exposure status to symptoms of postnatal distress (top); exposure status to at least one episode of symptoms of maternal distress (bottom). Unadjusted scores provide face validity to the adjusted growth curve models.

3.4.2 Specification of Underlying Growth Function

3.4.2.1 Specification of Time Metrics

There was large variation in the age at which children's SDQ measurements were obtained across each MCS sweep, as illustrated in Table 3.11, with age ranges for sweeps two

(eldest= 55 months) and three (youngest=53 months) overlapping. This justified the use of age at the time of SDQ assessment as the time metric rather than using discrete variables to signify which sweep measures were reported in. An additional quadratic time metric was specified as age^2. It was not possible to obtain children's age at GCA measurement as the GCA variable was derived from multiple test scores conducted at different points in time. Consequently, the growth curve models for GCA outcomes used sweep number as a discrete indicator for time.

Table 3.11: Summary Statistics of Children's Age (in months) at SDQ Measurement

	Mean	SD	Youngest	Eldest	Range
Sweep 2	37.65	2.46	32.02	55.00	22.98
Sweep 3	62.63	2.99	53.00	74.04	21.04
Sweep 4	86.76	2.96	76.01	98.07	22.06
Sweep 5	134.07	3.94	122.07	148.11	26.03

3.4.2.2 Transformation to SDQ outcome Variables

Square root transformations were applied for each SDQ outcome and compared to non-transformed outcomes in a random intercepts model (which included covariates for age, gender, ethnicity and family income). The most appropriate specification for the outcome variables were determined by assessing whether any modelling assumptions were violated i.e. if residuals were not normal or evenly distributed around the error terms. The square root transformation was adopted as it was found to be an improvement on the non-transformed outcomes: level 1 residuals deviated from the normal distribution for the non-transformed model, but not the square root transformed model. Diagnostics plots for the transformed/non-transformed SDQ total difficulties scale are illustrated in Figure 3.8.

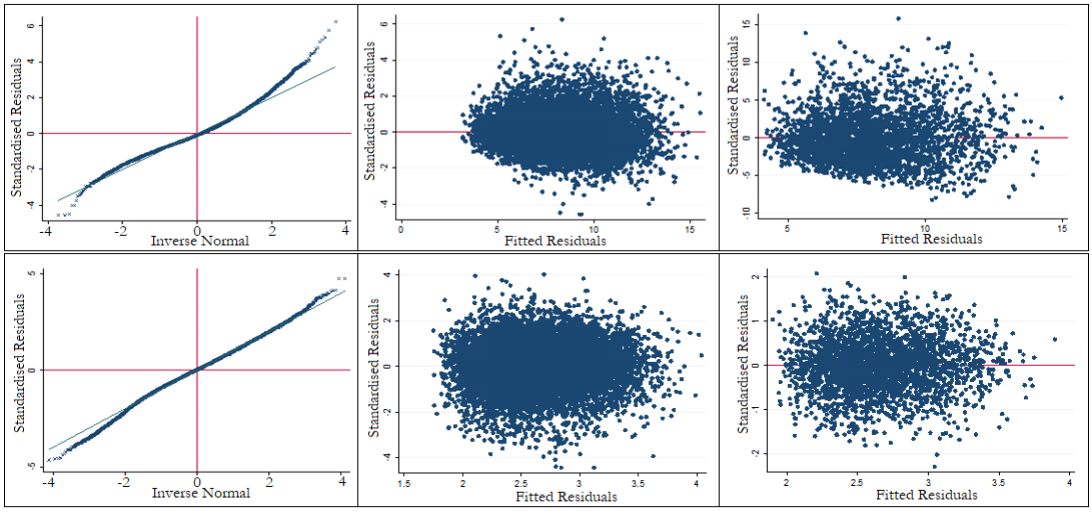


Figure 3.8: Residuals plots for random intercepts models comparing non-transformed (top) and square root transformed (bottom) SDQ total difficulties outcomes. Plots are provided for quantile normal plot (left) and scatter diagrams of standardised vs. fitted level 1 residuals (middle) and standardised vs. fitted level 2 residuals (right)

3.4.2.3 Growth Equations

As SDQ scores followed a curvilinear trajectory, models were fit with linear and quadratic terms for the time metrics (age and age^2). A model which included both linear and quadratic terms was found to significantly fit the data better than the model with only the linear term (Likelihood Ratio test $p<0.001$), Figure 3.9. Given that time metrics for the GCA outcomes were categorical the growth function was specified using dummy variables for sweeps 2-5 (i.e. whether the observation occurred at sweep x (yes/no)).

Next, random effects were added for each time metric. Random effects for age and age^2 (SDQ) and sweep number (GCA) were found to be significant ($p<0.001$) using likelihood ratio tests. The final underlying growth models for SDQ and GCA outcomes are described in Equation 3.1a & 3.1b. An unstructured covariance matrix was selected for each outcome variable as likelihood ratio tests suggested it to produce the best model fit when compared with first order autoregressive and exponential covariance structures.

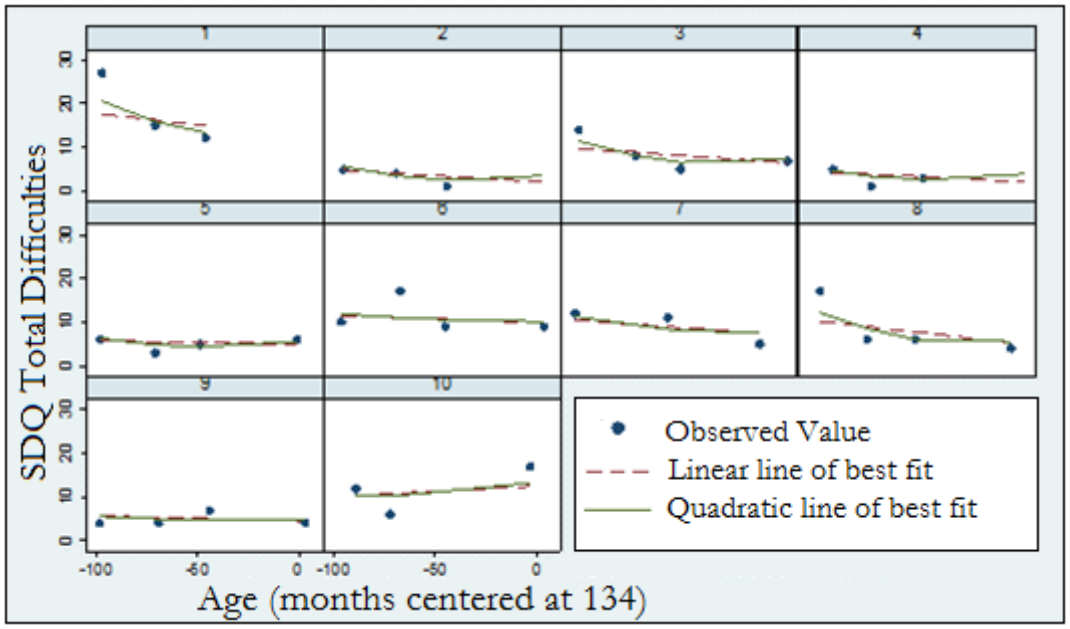


Figure 3.9: Predicted SDQ Total Difficulties Scores vs. age for 10 individuals for models with linear (red) and quadratic (green) terms for age.

Equation 3.1a: Underlying Growth Curve Specification for SDQ outcomes

$$Y^{1/2}_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + \beta_{2j}t_{ij}^2 + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j} \qquad \beta_{1j} = \beta_1 + u_{1j} \qquad \beta_{2j} = \beta_2 + u_{2j}$$

Equation 3.1b: Underlying Growth Curve Specification for GCA outcome

$$Y_{ij} = \beta_{1j}S2_{ij} + \beta_{2j}S3_{ij+} \beta_{3j}S4_{ij+} \beta_{4j}S5_{ij+} e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j} \qquad \beta_{1j} = \beta_1 + u_{1j} \qquad \beta_{2j} = \beta_2 + u_{2j} \qquad \beta_{3j} = \beta_2 + u_{3j} \qquad \beta_{4j} = \beta_2 + u_{4j}$$

For individuals:

$j = (1, .., n)$ with observations $i = (1, 2, 3, 4)$

Where:

Y=	SDQ or GCA scores
t=	age centred at Sweep 5 (age of child at SDQ measurement – 134 months)
S2=	Sweep two (yes/no)
S3=	Sweep three (yes/no)
S4=	Sweep four (yes/no)
S5=	Sweep five (yes/no)
e=	Level 1 residuals
u=	Level 2 residuals

3.4.3 Specification of Conditional Growth Curves

Conditional growth curves were estimated by adding covariates to the underlying growth function. Covariates were added one at a time and were all statistically significant ($p<0.05$) for the square root SDQ total difficulties outcome, while all covariates excluding the employment status of the mother were significantly associated with children's GCA score, Table 3.12.

Interaction terms were tested between each covariate and each model's time metrics, the significance of which is summarised in Table 3.12. Adding an interaction term between covariates and the time metric allowed the effect of the covariate to differ according to the age at which the observation occurred. In contrast covariate effect sizes remain constant across all sweeps where no interaction with time is specified.

Table 3.12: Explanatory Variables Included in SDQ and GCA Growth Curve Models

	SDQ		GCA	
	Covariate	Time Interaction	Covariate	Time Interaction
Time metrics				
Age	Yes	N/A	N/A	N/A
Age^2	Yes	N/A	N/A	N/A
Sweep Number	N/A	N/A	Yes	N/A
Maternal Depression Variables				
Postnatal Depression Symptoms	Yes	Yes	Yes	Yes
Maternal Distress Symptoms	Yes	Yes	Yes	Yes
Covariates				
Child Biological Sex	Yes	Yes	Yes	Yes
Child Ethnicity	Yes	Yes	Yes	Yes
Maternal Age at Child Birth	Yes	Yes	Yes	Yes
Number of Siblings	Yes	No	Yes	Yes
Household Income	Yes	Yes	Yes	Yes
Maternal Education	Yes	Yes	Yes	Yes
Single Mother	Yes	Yes	Yes	No
Mother Employment Status	Yes	No	No	No

Notes: Time interaction specified as an interaction term between the explanatory variable and the relevant time metric(s). All explanatory variables and time metrics were included in growth curve models if t-test results indicated the variable to be statistically significant ($p<0.05$).

3.4.4 Growth Curve Model 1 Results

The following results section focuses specifically on the effects of postnatal and maternal depression symptoms on children's development outcomes. The full regression results reporting coefficients for all covariates can be found in Appendix 3.3.

3.4.4.1 Postnatal Symptoms

Symptoms of maternal distress occurring during the postnatal period were significantly associated with increased (square root) SDQ total difficulties scores. Incremental differences between those exposed (n=1706)/ not exposed (n=10,098) to symptoms of distress *remained constant* for children aged three (0.38, $p<0.001$) to eleven (0.39, $p<0.001$), Figure 3.10 & Table 3.13. The age eleven exposure effect was estimated to equate to 2.07 on the SDQ raw scale following back transformation.

Children exposed to symptoms of maternal distress during the postnatal period were also associated with significant increases in both SDQ subscales: incremental differences between postnatal distress and (square root) SDQ internalising scores increased with each subsequent sweep with largest differences occurring at age eleven (0.32, $p<0.001$) (Figure 3.10); while (square root) SDQ externalising scores effect sizes diminished, but remained statistically

significant by age eleven (0.23, p<0.001). Interpretations remained consistent when comparing results for the imputed and non-imputed datasets for all outcomes, Table 3.13.

There were no statistically significant differences identified in GCA scores for those children exposed to postnatal symptoms (n=4682) vs. those not exposed (n=775). A null effect was observed across each of the four sweeps and for both the complete case and imputed databases, Table 3.13, Figure 3.10.

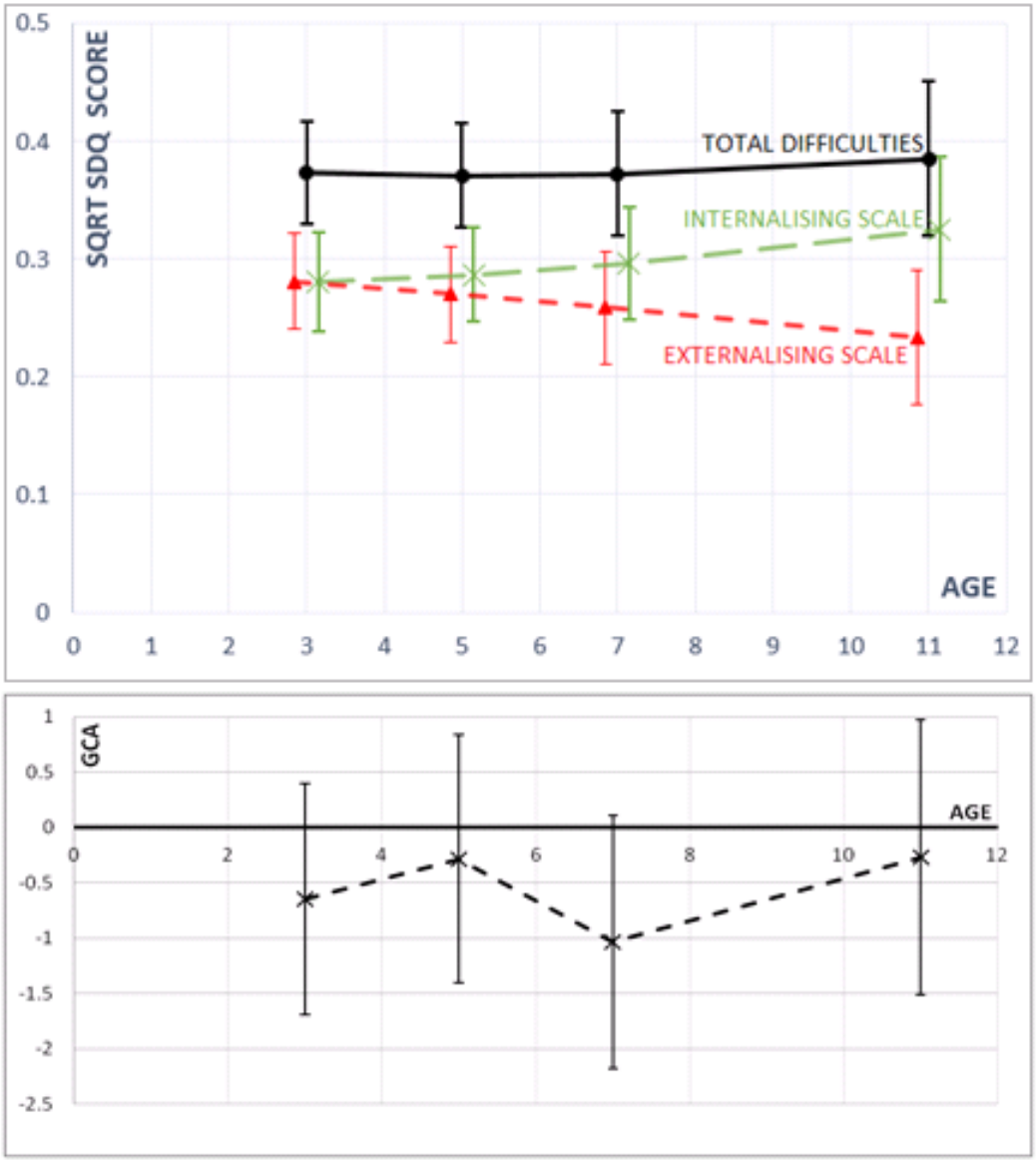


Figure 3.10: Incremental effect sizes and 95% confidence intervals from Growth Curve Model 1. Compares children exposed to symptoms of postnatal distress vs. non-exposed children for all square root transformed SDQ (top) and GCA (bottom) outcomes.

Table 3.13: Mean Incremental Effects in Children Exposed to PND vs. Not Exposed

	Coefficient	95% Lower CI	95% Upper CI	Pr>|z|
SQRT SDQ Total Difficulties				
Sweep 2	0.38	0.33	0.42	<0.0001
Sweep 3	0.37	0.33	0.42	<0.0001
Sweep 4	0.37	0.32	0.43	<0.0001
Sweep 5	0.39	0.32	0.45	<0.0001
Sweep 5 (Imputed)	*0.34*	*0.28*	*0.39*	*<0.0001*
GCA				
Sweep 2	-0.65	-1.69	0.40	0.2306
Sweep 3	-0.28	-1.40	0.84	0.5614
Sweep 4	-1.03	-2.18	0.11	0.0862
Sweep 5	-0.27	-1.51	0.98	0.6733
Sweep 5 (Imputed)	*-0.85*	*-2.00*	*0.29*	*0.1541*
SQRT SDQ Internalising Scale				
Sweep 2	0.28	0.24	0.32	<0.0001
Sweep 3	0.29	0.25	0.34	<0.0001
Sweep 4	0.30	0.25	0.34	<0.0001
Sweep 5	0.32	0.26	0.39	<0.0001
Sweep 5 (Imputed)	*0.29*	*0.24*	*0.35*	*<0.0001*
SQRT SDQ Externalising Scale				
Sweep 2	0.27	0.24	0.32	<0.0001
Sweep 3	0.27	0.22	0.31	<0.0001
Sweep 4	0.26	0.21	0.31	<0.0001
Sweep 5	0.23	0.18	0.29	<0.0001
Sweep 5 (Imputed)	*0.21*	*0.16*	*0.25*	*<0.0001*

Notes: Results obtained from Growth Curve Model 1 for the complete case database across all sweeps, and imputed databases (in *italics*) for Sweep 5.
SDQ (all scales) n=11,804 (*n=16,033)*; GCA n=5,457 *(6,643)*.

3.4.4.2 Accumulated Symptoms of Distress

A positive and statistically significant association was identified between (square root) SDQ total difficulties scores and accumulated symptoms of maternal distress occurring during sweeps 2-5. The coefficient attached to the cumulative distress variable decreased from 0.37 ($p<0.001$) at age three to 0.25 ($p<0.001$) for children aged eleven, Table 3.14. The age eleven cumulative distress effect size was equal to 1.19 on the raw SDQ total difficulties scale following back transformation.

Accumulated symptoms of distress were associated with reduced GCA scores and these were also statistically significant. Largest effects were observed at sweep two -2.05($p<0.05$) which persisted until sweep five -1.83 ($p<0.001$), Table 3.14. Results were consistent on both the SDQ externalising and internalising scales and all interpretations remained constant across complete case and imputed databases.

Table 3.14: Mean Effect per Episode of Maternal Distress Symptoms (Sweeps 2-5)

	Coefficient	95% Lower CI	95% Upper CI	Pr>|z|
SQRT SDQ Total Difficulties				
Sweep 2	0.37	0.29	0.44	<0.0001
Sweep 3	0.35	0.30	0.39	<0.0001
Sweep 4	0.32	0.27	0.36	<0.0001
Sweep 5	0.25	0.21	0.29	<0.0001
Sweep 5 (Imputed)	*0.26*	*0.23*	*0.29*	*<0.0001*
GCA				
Sweep 2	-2.05	-4.06	-0.05	0.0263
Sweep 3	-1.91	-3.16	-0.66	0.0024
Sweep 4	-1.59	-2.15	-0.11	0.0021
Sweep 5	-1.83	-2.62	-1.05	<0.0001
Sweep 5 (Imputed)	*-1.29*	*-1.99*	*-0.58*	*0.0005*
SQRT SDQ Internalising Scale				
Sweep 2	0.28	0.21	0.36	<0.0001
Sweep 3	0.30	0.25	0.34	<0.0001
Sweep 4	0.29	0.24	0.33	<0.0001
Sweep 5	0.22	0.18	0.27	<0.0001
Sweep 5 (Imputed)	*0.23*	*0.20*	*0.27*	*<0.0001*
SQRT SDQ Externalising Scale				
Sweep 2	0.27	0.21	0.34	<0.0001
Sweep 3	0.24	0.20	0.28	<0.0001
Sweep 4	0.21	0.17	0.25	<0.0001
Sweep 5	0.16	0.13	0.20	<0.0001
Sweep 5 (Imputed)	*0.17*	*0.14*	*0.20*	*<0.0001*

Notes: Table reports beta coefficient attached to the Cumulative Distress Variable, indicating the mean impact per episode of maternal depression for sweeps 2-5. Results obtained from Growth Curve Model 1 for the complete case database across all sweeps, and imputed databases (in *italics*) for Sweep 5.
SDQ (all scales) n=11,804 (*n=16,033)*; GCA n=5,457 *(6,643)*.

The effect of repeated exposures to maternal distress is illustrated in Figure 3.11, which depicts the predicted (square root) SDQ and GCA scores for Growth Curve Model 1 for a set of hypothetical trajectories of maternal distress symptoms. Predicted (square root) SDQ and GCA scores are affected most largely for children whose mothers suffer from persistent distress (symptoms at sweeps one, two, three, four and five), then episodic symptoms (arbitrarily defined as symptoms at sweeps one, two and four) and lastly for children whose mothers suffer from distress only during the postnatal period (sweep one).

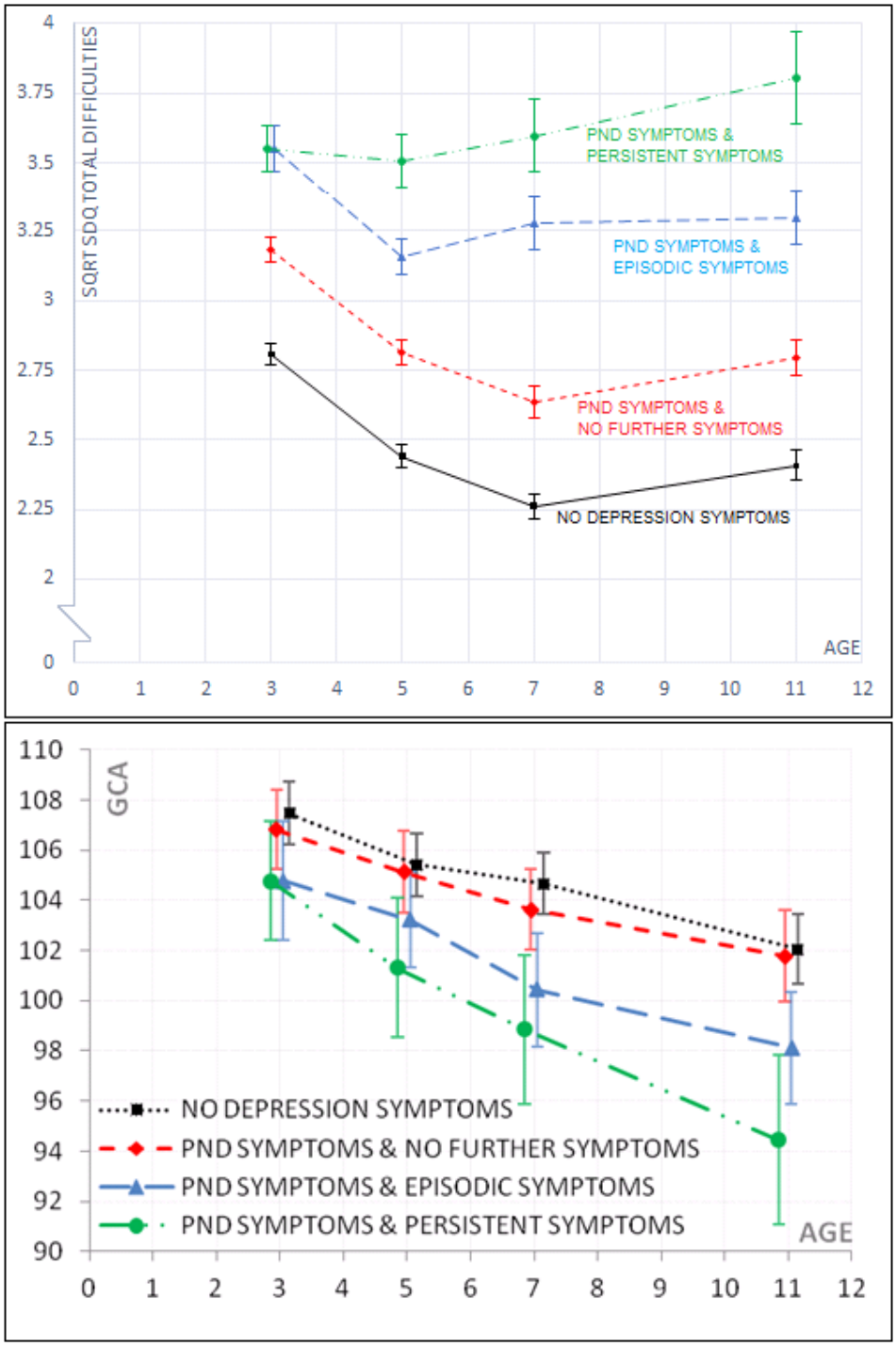


Figure 3.11*:* Plots the predictions of Growth Curve Model 1 for four hypothetical individuals for sqrt SDQ total difficulties (top) and GCA (bottom). Predictions are obtained for children whose mothers suffer from (1) no symptoms of postnatal or maternal distress, (2) symptoms of postnatal distress but not maternal distress, (3) symptoms of postnatal distress and episodic symptoms of maternal distress which occur at sweeps 2 & 4 (but no symptoms at sweeps 3 & 5), and (4) symptoms of postnatal distress and persistent symptoms of maternal distress at sweeps 2, 3, 4 and 5. Each individual is assumed to have the following reference case covariate characteristics: child gender= female; child ethnicity= white; child siblings at birth=0; mother's qualification= NVQ level 2, mother's income= not below 60% of median, mother's age= 28, mother's relationship status=married; mother's work status= not in work).

3.4.5 Growth Curve Model 2 Results

Growth Curve Model 2 grouped children according to their mother's trajectory of symptoms through sweeps 1-5. To further investigate the effects of PND three pairwise comparisons were made between mothers' distress trajectories. Comparisons were made between trajectories which reported the same frequency of symptoms after the postnatal period but differed according to PND symptoms:

1. No PND with no further symptoms (SDQ n=6548/GCA n=3516) vs. PND with no further symptoms (SDQ n=736/ GCA=383)
2. No PND with episodic maternal symptoms (n=372/203) vs. PND with episodic maternal symptoms (n=229/120)
3. No PND with persistent maternal symptoms (n=23/15) vs. PND with persistent maternal symptoms (n=62/34)

The effect of PND symptoms on (sqrt) SDQ total difficulties scores at sweep five appeared to be dependent on the presence of later depression: if no later depression occurred then PND had a significant impact (0.37, $p<0.01$); there was no significant difference between children whose mothers had later episodic depression with or without PND (0.11 $p>0.05$); while children whose mothers had persistent depression with PND scored lower than those without PND (-0.04, $p>0.05$), Figure 3.12. No effects were identified in any of the pairwise comparisons for the GCA outcomes where each incremental effect size was not significantly different to zero ($p<0.05$).

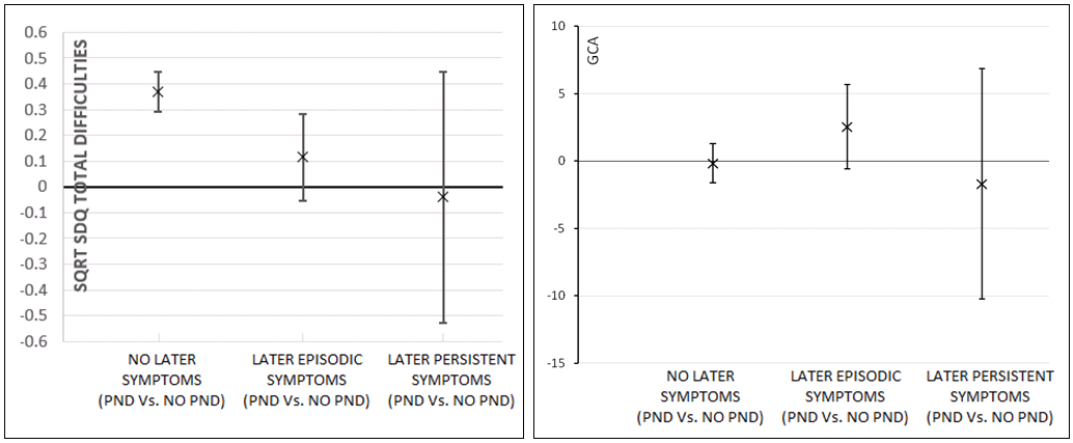


Figure 3.12: Incremental effect sizes identified in pairwise comparisons between maternal distresses trajectories. Later episodic symptoms were defined as 1-2 episodes of distress between sweeps 2-5; later persistent symptoms were defined as 3-4 episodes of distress between sweeps 2-5. Results are obtained at sweep five (age eleven) for square root SDQ Total Difficulties (left) and GCA (right).

As with results from Growth Curve 1, when compared with groups who had no symptoms of depression between sweeps 2-5, substantial effects were identified for children exposed to episodic and persistent symptoms. These results are presented in full in Appendix 3.4.

3.5 Discussion

3.5.1 Interpretation of Results

3.5.1.1 Key Findings

The results from this study suggest that symptoms of maternal depression may affect child development, at least up to the age of eleven. Symptoms of PND were significantly associated with socioemotional but not cognitive development outcomes when adjusting for symptoms of maternal distress occurring later in childhood. The study provided no evidence of increasing/diminishing effects for PND on child development, rather the incremental effect sizes for both SDQ and GCA outcomes remained consistent across all sweeps.

3.5.1.2 Latency Model

There was no evidence suggesting that the immediate postnatal period is specifically critical for the effects of maternal depression on child development. Symptoms of PND were associated with increased internalising, externalising and global socio-emotional problems, but effects were not distinct from the effects of symptoms occurring during later sweeps. Furthermore, negligible differences were identified for children exposed/not exposed to symptoms of PND if their mothers suffered from later episodic or persistent depression.

No association was found between PND and children's cognitive development at any age. Interestingly, PND was initially associated with detrimental effects to GCA (solely for a latency model), but these were attenuated by later cumulative distress when this variable was added to Growth Curve Model 1. These results demonstrate the importance of adjusting for later periods of depression when attempting to identify effects occurring specifically during the postnatal period.

The results from this study are largely consistent with the existing literature. In a systematic review Kingston and Tough (2014) concluded that, between ages 2-8, socioemotional but not cognitive development was likely to be affected by symptoms of postnatal depression, finding no evidence for increased sensitivity during the postnatal period. Meanwhile Sanger et al. (2015) reviewed longitudinal studies that reported (non-significant) associations between postnatal depression and increased internalising and externalising symptoms in adolescents aged 11-16.

The increased precision resulting from the larger sample size used here may explain why these results were statistically significant.

Inconsistencies were identified with studies by Murray et al. (2011) and Hay et al. (2008) who found an association between symptoms of postnatal depression and cognitive/academic outcomes in adolescence. Differences in study design may explain these discrepancies. For instance, PND symptoms were measured here at nine months following birth compared with at six weeks (Murray et al., 2011) and three months (Hay et al., 2008). Furthermore Murray et al. (2011) did not account for maternal depression after the postnatal period and both studies used diagnosis of depression rather than symptoms of distress as the postnatal exposure.

3.5.1.3 Accumulation of Symptoms

This study provides robust evidence in support of the accumulation hypothesis. A dose dependent relationship was identified between the number of times children were exposed to symptoms of maternal distress and worse scores on measures of both cognitive and socioemotional outcomes. These results were replicated when mothers were grouped according to the frequency of their depression symptoms, with persistent symptoms associated with larger effect sizes than episodic symptoms.

These results extend findings by Brennan et al. (2000) and Kingston and Tough (2014) who identified significant effects for chronic maternal depression on early child development. The results also provide further vindication for conclusions drawn by Sanger et al. (2015) who, based on studies with limited sample sizes, suggested that persistent maternal depression was likely to be associated with adverse cognitive and socioemotional outcomes in adolescence.

The effect of cumulative distress on cognitive and socioemotional outcomes appeared to be largest for observations at sweep two and diminished for each subsequent sweep. This might be interpreted as evidence that early periods were more sensitive to the effects of maternal distress. Alternatively, the reduced marginal effect of cumulative distress during later sweeps may have occurred as a result of a dose-dependent response plateauing with each increasing episode. That is, the impact of moving from three to four instances of symptoms is less than moving from a frequency of one to two. Therefore, the cumulative distress coefficient may have been larger in earlier sweeps given the lower maximum frequency of symptoms associated with these sweeps. This latter interpretation is supported by results from Growth Curve Model 2 which identified differences between PND and no PND which decreased in size with the increased frequency of later symptoms.

3.5.2 Policy Implications

The negative impact of maternal depression on child development is likely to have implications for health policy. Measures of socioemotional development such as the SDQ can provide an indication of children's mental health status, for example, Furber et al. (2014) suggest that SDQ scores represent child utility and map SDQ onto the Children's Health Utility-9 scale in an Australian sample. Additionally, and as outlined in chapter two, increasing evidence links child development with prosperity during adulthood across health (Campbell et al., 2014), education (Rabiner et al., 2016), crime (Frisell et al., 2012) and employment (Heckman, 2006). Health technologies that prevent the exposure of children to maternal depression symptoms, therefore, have the potential to be cost-effective by both improving early life health whilst also reducing the detrimental effects of PND on children's later lifespan development. Typical treatments for depression include psychological and pharmacological therapy and there is evidence from several randomised studies suggesting that both can significantly reduce depression symptomatology in postnatal women (Cuijpers et al., 2008).

There is also promising evidence suggesting treatment benefits for maternal symptoms may translate into benefits on child outcomes. In a meta-analysis of randomised studies Cuijpers et al. (2015) identified the effects of psychological therapy for PND versus a relevant control (e.g. usual care) on *both* maternal and child outcomes. The meta-analysis by Cuijpers et al. (2015) included nine studies measuring treatment effects on maternal depression outcomes, seven of which also measured effects on behavioural, emotional or mental health screening questionnaires for children between ages of nine months and two years. Results by Cuijpers et al. (2015) identified significant reductions in maternal depression symptoms following psychological therapy (standardised effect size g=0.66) and significant reductions on children's mental health symptoms where effects sizes were around two thirds (g=0.40) of those observed on maternal outcomes.

The translation of treatment effect across maternal and child outcomes is likely to be mediated through the mother-child attachment relationship, thus providing an alternative target for PND intervention. The meta-analysis by Cuijpers et al. (2015) included a third analysis which identified a significant effect for psychological therapy (vs. control) on outcomes measuring the quality of mother-child interactions. Effect sizes (g=0.38) were almost equivalent to the effect size on child outcomes (g=0.40). Parenting interventions target the maternal-child attachment relationship and there is some empirical evidence identifying positive effects for these interventions on toddlers' emotional and behavioural development outcomes (Fitelson et al., 2011). Treatment for depression and parenting interventions need not be delivered in isolation. As suggested by Goodman (2004) the most effective intervention strategy for reducing the

effects of postnatal depression on child development might be an integrated approach that offers psychological/pharmacological treatment for postnatal depression *alongside* parenting interventions.

While there is evidence demonstrating the benefits of intervention during the postnatal period, the results from this study suggest that children are at most risk if they are exposed to accumulated symptoms over time. The largest treatment benefits are likely to be achieved for health technologies which target chronic maternal depression occurring throughout child development. It is the view of Goodman and Garber (2017) that the benefits of parenting interventions could be maintained across childhood through annual visits which provide mothers with skills relevant to the child's developmental stage. Similarly, large treatment benefits could be achieved through pharmacological/psychological therapy if these reduce the probability of depression reoccurrence. There is currently a sparsity of randomised literature identifying the long-term effects of depression treatments/parenting interventions and this should be an objective for future research.

Finally, health technologies which screen women for postnatal depression soon after birth could enable the timely delivery of treatments and may enhance the early child attachment relationship and prevent detrimental exposure to maternal depression symptoms. Current guidelines regarding the cost-effectiveness of screening for postnatal depression do not formally account for effects of depression on children (NICE, 2018). The cost-effectiveness of this guidance is reassessed in an economic evaluation in chapter six using evidence from this chapter to inform some of the model's parameters.

3.5.3 Methodological Considerations for Indirect Estimation

With respect to the overarching aim of the thesis, this chapter demonstrates methodologies appropriate for the first stage of indirect estimation where the effects of an early life circumstance (exposure to PND symptoms) were estimated on intermediate lifespan development endpoints. This empirical analysis could inform the design of future research as it addresses three methodological challenges likely to be faced by other researchers whose objective is to (indirectly) identify the lifetime effects associated with early life circumstances through effects to lifespan development.

Firstly, this analysis accounts for potential temporal changes in lifespan development effects sizes by estimating growth curve models in time series data. The use of growth curve modelling is desirable for indirect estimation as it can provide evidence to justify/dispute assumptions regarding the continuation of the indirect effect pathway (from early life circumstances to intermediate child development measures and through to lifetime decision endpoints). The

results of this analysis identified consistent effect sizes between postnatal depression symptoms and both cognitive and socioemotional endpoints between the ages of three and eleven, which might suggest a persistent effect for postnatal depression on child development outcomes. Therefore, it might be reasonable to assume that the detrimental effects of postnatal depression symptoms continue beyond the study time horizon, into adolescence and further into adulthood. It would not have been possible to obtain this information from traditional study designs which identify effect sizes at a single follow up period.

Secondly, there is a question regarding the selection of outcome measures which can appropriately act as intermediate lifespan development endpoints for indirect estimation. Because development is such a vast process, no single measure is available, instead analysts rely on measures specific to the cognitive, socioemotional and/or physical domains. Researchers should search the literature to establish which domains are most likely to be affected by the specific early life circumstance under investigation. Both theoretical and empirical literature links cognitive and socioemotional development to maternal postnatal depression (e.g. (Bowlby, 1978), (Brennan et al., 2000), (Sanger et al., 2015)), which led to the investigation of these domains in this analysis.

This analysis adopted global measures to characterise all the lifespan development effects within the cognitive and socioemotional domains. As informed through the psychometric literature GCA is a cognitive construct that underlies performance on narrower skills, whilst the SDQ total difficulties scale combines scores for both emotional and behavioural problems. In general, it is important for global development measures to be adopted as intermediate endpoints for indirect estimation as a focus on more specific abilities could risk underestimating effect sizes. This risk is illustrated by comparing results of the different SDQ outcomes for Growth Curve Model 1 where the detrimental effects associated with postnatal depression were lower in magnitude for the narrower Externalising and Internalising subscales when compared with the Total Difficulties scale. Methods in this analysis, which used PCA to identify GCA might be applicable to other research settings as it is not uncommon for longitudinal cohort studies to administer incomplete psychometric scales or apply different types of cognitive tests across sweeps as in the MCS.

Thirdly, the results from this chapter demonstrate the complicated pathways of effect which can occur between early life circumstances and lifespan development outcomes. Inappropriate specifications of empirical models could result in the identification of different effect sizes. For example, the effects of postnatal depression may have been overestimated if it had been assumed that the pathway of effect was solely accounted for by a latency model. As was previously described the effect of PND symptoms on cognitive outcomes were found to be completely attenuated by the addition of accumulated maternal depression symptoms after the

postnatal period. Analysts should consult the existing evidence base to identify the most plausible pathway of effect which is likely to be complex and specific to the early life circumstance under investigation.

The methods applied in this analysis could be extended by accounting for the interacting effects which may occur between cognitive and socioemotional development outcomes. Multivariate growth curve models include more than one outcome variable within the same model (Curran et al., 2012) and could simultaneously identify GCA and SDQ development trajectories. By incorporating more complex error structures multivariate models can establish the covariance between domain specific outcomes (Curran et al., 2012), capturing the cross productive effects between outcomes in different domains (i.e. the influence of cognitive development on socioemotional development and vice versa).

A multivariate model which incorporates three intermediate lifespan development outcomes with global measures in each of the cognitive, socioemotional and physical domains would provide the most complete approach when identifying the developmental effects associated with early life circumstances for indirect estimation. Such a model would be desirable as it could account for the holistic nature of lifespan development.

3.5.4 Limitations

There were several limitations regarding the measures used in this study. Firstly, the 9-item Rutter malaise instrument and the Kessler-6 Index were used to measure symptoms of maternal distress and are not diagnostic instruments. Therefore, this study could only confirm that mental health *symptoms* rather than *diagnoses* were associated with child development. As similar results have been obtained on instruments with diagnostic validity, such as the Edinburgh Post Natal Depression Scale (Sanger et al., 2015), it might be reasonable to assume that effects of depression symptoms translate to those with a depression diagnosis.

Secondly, interpreting differences in effect sizes between sweep one and sweeps 2-5 was problematic given that different instruments with different scales were used to identify distress symptoms. The validated threshold for the Kessler 6 index (>=13) identifies cases of *severe* maternal distress (Prochaska et al., 2012), while the 9-item Rutter malaise index identifies *moderate* symptoms (Dex and Joshi, 2004). Results were obtained for a lower threshold value of the Kessler 6 Index (>=5) which, as suggested by Prochaska et al. (2012) is a non-validated indicator of *moderate* symptoms. Interpretation results remained consistent across the higher and lower thresholds, results for the lower threshold model are reported in Appendix 3.5

Thirdly, this study used maternal reports for SDQ outcomes which may have resulted in reporter bias if depressed mothers over- reported their child's socioemotional problems. Some discrepancies have been identified between maternal/non-maternal reported outcomes (Leis et al., 2014). However, in their review Leis et al. (2014) find evidence that effects of maternal depression on socioemotional development occur independently of reporting errors, and Kingston and Tough (2014) find that significant effects remain in most studies that use non-maternal report. Leis et al. (2014) even suggest that depressed mother might be more accurate reporters than non-depressed mothers according to the *depressive realism* hypothesis.

Finally, as this was not an experimental study a causal relationship between maternal distress symptoms and child development cannot be confirmed. It is possible that omitted confounding factors explain the observed relationships, for instance: antenatal depression is associated with an increased risk of PND and is also associated with child development (Kingston and Tough, 2014), (Sanger et al., 2015). It is unlikely that the entire effect sizes identified in this analysis are explained by antenatal depression as several studies have established effects for maternal depression before and after birth on global child development measures within the same model (Kingston and Tough, 2014). Similarly, paternal PND is correlated with maternal PND and detrimentally associated with child development outcomes (Ramchandani et al., 2005). However, evidence from Nuttall et al. (2019) suggests the effects of paternal symptoms may occur indirectly through their influence on maternal symptoms. Thus, the inclusion of paternal symptoms in an interdependent model might enhance (rather than confound) the effects of maternal depression.

3.5.5 Conclusions

When placing the results from this longitudinal analysis within the context of existing research there is compelling evidence linking maternal depression with effects to child socioemotional and cognitive development. This study identified an association between postnatal depression symptoms and child development; however, there was no evidence to suggest that the immediate post-natal period was critical. Rather, larger effects were identified if children were exposed to repeated symptoms that may have accumulated throughout their early development. Children are likely to be at highest risk if their mother suffers from postnatal depression and develops later chronic depression. Screening for maternal depression soon after birth could provide an early opportunity for intervention which may reduce the exposure of children to harmful depressive symptoms.

The methods used in this chapter inform the first stage of a larger methodology estimating the lifetime effects associated with postnatal depression. This empirical analysis could be used to

inform the design of future research which attempts to establish the lifetime effects associated with other early life circumstances.

3.6 Acknowledgements

Data from the Millennium Cohort Study (UoL, 2017) was supplied through the UK Data service archive at the University of Essex. Secure Access to the KS1 (UoL, 2015a) & KS2 (UoL., 2015b) Linked Education Administrative Datasets required the author to obtain ESRC Accredited Researcher status from the UK Data Service. Statistical Disclosure Control has been undertaken by the UK Data Service on all Secure Access data outputs to ensure that results cannot be used to identify individual respondents. All results and interpretations are the view of the author.

Chapter 4: A Scoping Review Identifying Mathematical Models Predicting the Lifetime Health and Economic Outcomes of Child Development

4.1 Summary

The analysis in the previous chapter addressed the first stage of indirect estimation and identified the effects of postnatal depression on intermediate endpoints for children's cognitive and socio-emotional development. The current chapter focuses on the second stage of indirect estimation which requires effects on lifetime decision endpoints to be predicted from the previously identified effects on intermediate development measures. Mathematical modelling is often used in economic forecasting and could be appropriate when forecasting lifetime effects if models can appropriately characterise the relationship between cognitive and socio-emotional outcomes and the relevant decision endpoints for economic evaluation. This chapter describes a scoping review aiming to identify an appropriate mathematical model from existing literature. Upon failing to identify any fully relevant studies the search criteria of the review are extended, and a second aim is devised, namely, to inform the design of future research (in chapter five) by identifying key modelling characteristics through a thematic analysis. Figure 4.1 depicts the contribution of evidence in this chapter towards the overall estimation of lifetime effects in this thesis.

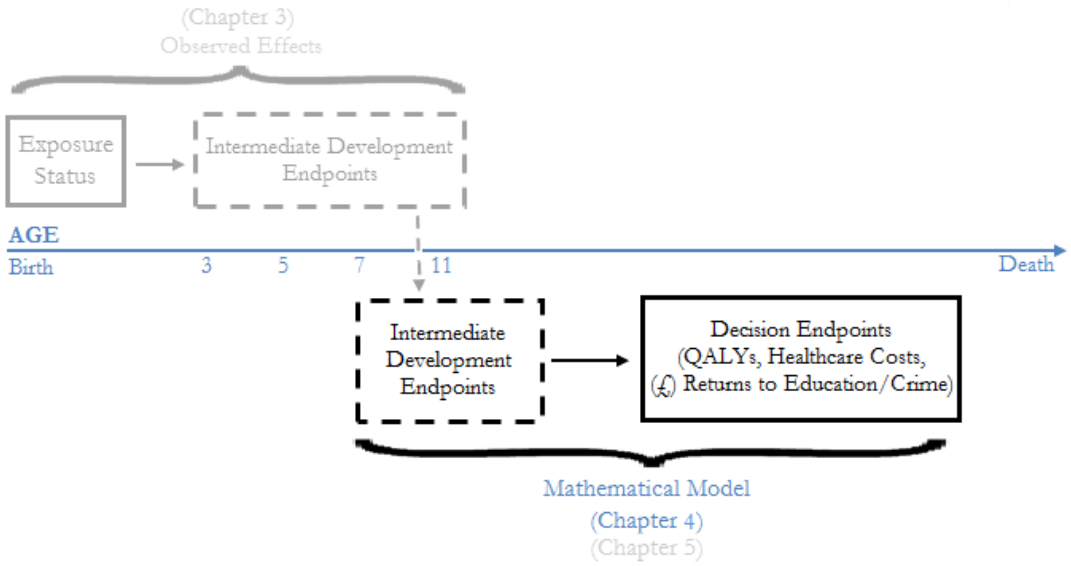


Figure 4.1: Demonstrates how this chapter contributes to the overall estimation of lifetime effects.

4.2 Introduction

4.2.1 Quantitative and Qualitative Forecasting

Indirectly estimating the lifetime effects of early life circumstances requires an appropriate method to predict how unit changes on intermediate outcomes translate to unit changes on lifetime decision endpoints. Specifically, the applied example in this thesis requires predictions of expected Quality Adjusted Life Years (QALYs), lifetime healthcare costs, monetary costs/savings in government sectors outside of health, and economic productivity effects from unit changes in children's Strength and Difficulties Questionnaire (SDQ) and General Cognitive Ability (GCA) scores. In the economics literature, predictions of this type are often referred to as economic forecasts.

In general, forecasting can be approached from either a quantitative or qualitative basis (Bowerman et al., 2005), (Makridakis et al., 2008). Quantitative forecasts make future predictions based on the analysis of historical data, whereas qualitative forecasts are identified from the subjective opinions of individuals, usually experts (Bowerman et al., 2005). Due to their subjective nature, the reliability of qualitative forecasts can be questioned. Makridakis et al. (2008) suggest that qualitative forecasts could be useful when predicting a range of possibilities but are less useful when identifying point estimates and should only be used for this purpose if quantitative data is scarce or unavailable.

Quantitative forecasting is selected as the more applicable forecasting approach in this research as the ultimate purpose for estimating lifetime effects is to inform parameter estimates in an applied economic evaluation – which is itself a predominantly quantitative exercise.

4.2.2 Mathematical Models for Forecasting in Economic Evaluation

Mathematical models in their broadest sense are descriptions of systems using mathematical language and concepts and are therefore useful tools for obtaining quantitative forecasts. A large variety of models exist; the most common types used in economic forecasting include decomposition models, dynamic models, exponential smoothing models, linear regression models and time series models (Bowerman et al., 2005), (Makridakis et al., 2008).

Mathematical models could provide an appropriate method when forecasting effects from intermediate endpoints to final decision endpoints in economic evaluation (Brennan and Akehurst, 2000), (Buxton et al., 1997), (Garnett et al., 2011). Examples of mathematical models applied to forecasting lifetime outcomes in economic evaluation include the UKPDS Outcomes simulation model which is used to estimates lifetime QALYs and healthcare costs by modelling

the pathway between multiple earlier risk factors and the onset of diabetes (Clarke et al., 2004); a model by Oster et al. (1999) which uses logistic regression models to account for the pathway between cholesterol and coronary heart disease; and the National Heart Formulation model which simulates the lifetime BMI related healthcare costs for a hypothetical cohort of 50,000 children (McPherson, 2007) and is used by Hollingworth et al. (2012) when assessing the cost-effectiveness of lifestyle interventions to treat overweight and obesity in children.

Appropriate models should characterise the different diseases, processes and relationships occurring between intermediate and final decision endpoint as illustrated in Figure 4.2. There is likely to be a complicated relationship between intermediate developmental outcomes and lifetime decision endpoints: adulthood outcomes such as health and employment are influenced by lifestyle factors (e.g. smoking, diet, exercise etc.) and capabilities (e.g. cognitive function, personality, motivation). Factors and capabilities are determined by the lifespan development process which continues throughout adolescence and early adulthood.[4] An appropriate mathematical model used to forecast lifetime effects for indirect estimation should account for these processes between intermediate outcome and final decision endpoint.

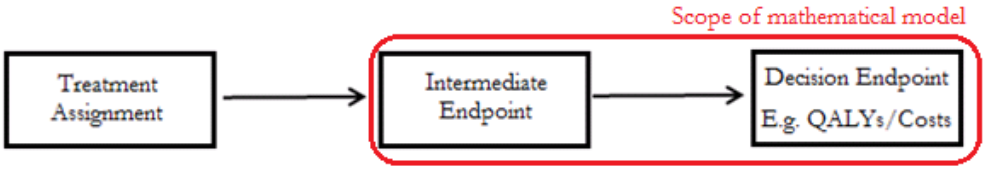


Figure 4.2: Illustrates the path from indirect estimation of treatment effect on decision endpoint via a mathematical model. The modelling process occurs within the red box where equations are used to establish the pathway and relationship between model inputs (intermediate development endpoints) and model outcomes (decision endpoints). Figure is adapted from Garnett et al. (2011) (see Figure 1 pp. 516).

4.2.3 Research Objectives

The purpose of this chapter is to establish an approach for the second stage of indirect estimation (see figure 4.1 above). This allows the estimation of the incremental lifetime effects of children's exposure to postnatal depression to be predicted from the intermediate effects identified on developmental outcomes in chapter three.

The primary objective of the research described in this chapter was therefore:

See chapter two sections 2.2.5 & 2.2.6.

> To identify a mathematical model in the existing literature that could be used to extrapolate the previously observed effects of postnatal depression on children's cognitive and socio-emotional development at age eleven (see chapter 3) to lifetime decision endpoints.

As will be reported later in this chapter, no suitable mathematical model could be identified in the existing literature. Therefore, additional primary research was required to provide the data needed to support the indirect estimation of lifetime effects in this thesis. This extra evidence is obtained through mathematical modelling, the results of which are reported in chapter five. Consequently, a second objective for the current chapter was devised to enable the design of the further primary analysis, namely:

> To conduct a thematic analysis to identify key modelling characteristics that could be applied to the design of a primary analysis used to forecast lifetime effects from intermediate development outcomes.

4.3 Methods

4.3.1 Analytical Approach

A scoping review was chosen as the method of enquiry to achieve the research objectives. Scoping reviews are commonly used to identify the extent of and map literature for novel research questions. This contrasts with systematic reviews which are typically used to synthesise evidence for more specific research questions e.g. treatment effectiveness, specificity/sensitivity of diagnostic tests, prevalence of disease etc. (Arksey and O'Malley, 2005), (Levac et al., 2010).

To achieve the first objective, the inclusion criteria for the scoping review specifically required the estimation of the decision endpoints relevant for economic evaluation i.e. QALYs, healthcare costs, cross-sectoral costs and economic productivity. To achieve the second objective, the inclusion criteria were relaxed to include models with outcomes that are not the specific decision endpoint of interest for economic evaluation i.e. not necessarily QALYs and/or healthcare costs but additional clinical or economic endpoints including outcomes such as blood pressure, life-satisfaction, and high school graduation success. The scoping review therefore included a thematic analysis to identify key modelling characteristics across studies identified through the relaxed search criteria.

The methods used in this chapter followed recommendations in publications by Arksey and O'Malley (2005), Colquhoun et al. (2014) and Levac et al. (2010). These studies appraise the

methods used for scoping studies and are often cited in published reviews, for instance in O'Flaherty and Phillips (2015) and Pham et al. (2014).

4.3.2 General Iterative Strategy

Unlike systematic reviews, the search strategy used in scoping reviews can be flexible and iterative. This means researchers can reflect upon and repeat search and selection processes before defining final eligibility criteria (Arksey and O'Malley, 2005), (Levac et al., 2010). This is particularly useful for researchers with novel objectives working within a large literature base as strategies can be conducted and refined within subsections of the literature. The final eligibility criteria for this review were derived following several iterative searches carried out within a single electronic database.

The general selection strategy was to (i) propose a broad set of inclusion criteria from which a search strategy was derived, (ii) apply the search across a single database and (iii) review and refine the search strategy over several iterations until a manageable number of results were returned. The next phase was to (iv) pilot the broad eligibility criteria across a small number of studies (v) review the returned results and (vi) restrict the eligibility criteria if the proportion of included studies appeared to be unmanageable. This general strategy is shown in Figure 4.3.

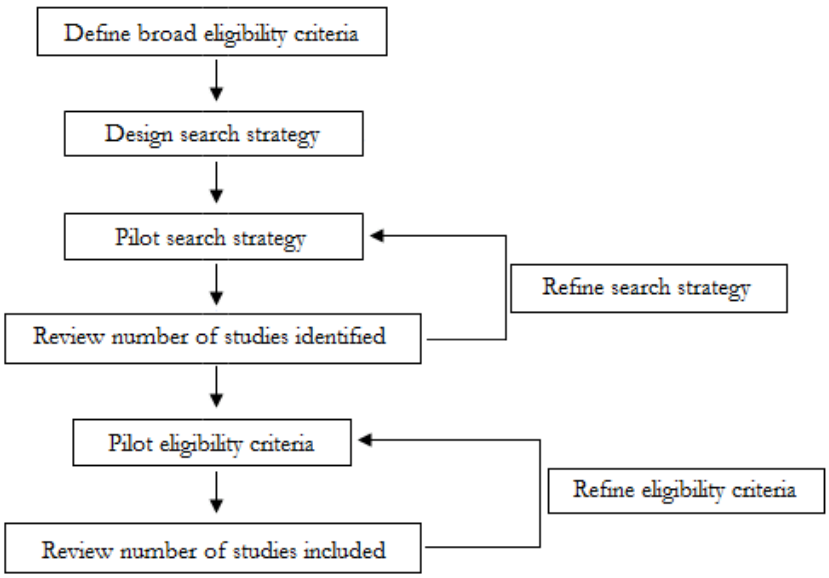


Figure 4.3: Illustrates the iterative methods used for devising the final search strategy and eligibility criteria

4.3.3 Eligibility Criteria

For quantitative systematic reviews the acronym PICO (Population, Intervention, Comparator, and Outcome) is often used to formulate inclusion and exclusion criteria and design search strategies (Cooke et al., 2012). Some of these terms can be irrelevant in scoping reviews, for example there may be no intervention or comparator being investigated, as is the case in this analysis. Cooke et al. (2012) recommends an alternative acronym SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, and Research type) for use in scoping reviews. This analysis developed eligibility criteria according to the SPIDER acronym, details of which are provided below:

Study sample: The research objectives required a model which could extrapolate intermediate outcomes in mid childhood (around the age of 11) to adulthood. Therefore, the study sample was restricted based on the time horizon of the model. Studies were only included if the model contained variables for children aged 12 or younger. This was chosen as a cut-off point for mid-childhood given that it is usually the maximum age for primary education for children in the UK. Studies were also excluded if they did not obtain an outcome during adulthood which was defined as age 16 or above. This was chosen as it is the earliest age that individuals can leave education and enter employment in the UK.

Phenomenon of Interest: The review attempted to identify models that captured the effects of child development on later life outcomes. Therefore, it was necessary that included studies were informed by theoretical literature on child/lifespan development. No restrictions were placed on the discipline of the literature (e.g. psychology, health, social sciences etc.). Additionally, studies were excluded if their outcomes were restricted to a single disease as they were likely to provide models of specific disease processes that may not appropriately model development for a general population without the specific disease.

Design: Studies were excluded if their models were not estimated using real data. This meant that all empirical models based on primary observation (e.g. statistical models) were included, as well as any evidence synthesis models which combined multiple sources of secondary data. Purely theoretical models were excluded.

Evaluation: Studies were included/excluded according to the outcomes estimated by the models. The primary objective specifically required models to estimate QALYs and/or healthcare costs as these are key decision endpoints required for economic evaluation conducted from both a health and cross-sectoral decision maker's perspective.[5] The secondary

[5] For discussion of health and cross-sectoral decision maker's perspective see Chapter 2 sections 2.2.3 and 2.2.4.

objective relaxed the requirements by including studies that predicted outcomes that were relevant clinical or economic endpoints. Relevant clinical endpoints were defined as those which could either be used to compile QALYs (i.e. mortality/life-years), be used as substitutes/surrogates for QALYs/healthcare costs (e.g. blood pressure, BMI), or could potentially be mapped onto QALYs (e.g. self-reported measures of wellbeing/life satisfaction). Relevant economic endpoints were defined as being either preference-based measures, monetary outcomes, or outcomes which could be monetised through secondary literature (e.g. number of violent crimes committed, which might be monetised by estimating the mean cost per violent crime).

Research Type: Due to the size of the evidence base, studies were limited to research written in English and published after the year 2000. All animal-based studies were excluded.

4.3.4 Literature Sources and Search Design

The search strategy was aided by discussions with a qualified librarian, who provided guidance on appropriate literature sources, search terms, synonyms and subject headings. In total 6 electronic databases were searched: MEDLINE, PSYCINFO, ECONLIT, EMBASE, MATERNITY & INFANT CARE (Ovid), and CHILD DEVELOPMENT & ADOLESCENT STUDIES (Ebscohost). The search included studies published between 1st January 2000 and 31st August 2016.

The search design was based on the broad eligibility criteria (SPIDER) described above. Key search terms with their associated SPIDER category were: (1) "Child"- ***S****ample*; (2) "Development" – ***P****henomenon of* ***I****nterest*; (3) "Model" – ***D****esign*; (4) "Economic" – ***E****valuation.*

Multiple synonyms were included for all key search terms which were obtained from thesauruses, similar published literature, and using electronic database subject headings such as MeSH terms on MEDLINE. Economic outcomes were identified using published search filters designed to locate economic studies specifically from the electronic databases MEDLINE, EMBASE and PSYCINFO (Glanville et al., 2008). If no specific search filter was available common economic terms were derived from all other filters. Iterative searches were conducted to determine adjacency operators that optimised search specificity and sensitivity. The final search strategy is summarised below and is described in full in Appendix 4.1.

1. "Child", "development" attached with an adjacency operator (n=2)
2. "Child development" as a subject heading
3. Specific model types e.g. "mathematical", "probabilistic" attached to "model" with an adjacency operator (n=5)

4. Specific modelling features (e.g. "output" or "outcome" or "input" or "variable" or "function"), attached with adjacency operator (n=5)
5. "Model" as a subject heading
6. Study filter for "economic" studies
7. Additional economic outcome terms e.g. "cost", "human capital" "QALY"
8. 1 or 2
9. 3 or 4 or 5
10. 6 or 7
11. 8 and 9 and 10

The results of the final search strategy for each of the 6 databases were imported into the citation manager Endnote 7.0. Due to time constraints the strategy did not involve hand searching journals, searching grey literature or looking through reference lists. The extent to which this limits the review is discussed later. The full search and screen strategies were conducted by the author. Regular meetings were held with two senior researchers throughout all stages of the review. This allowed for discussion of any methodological issues that were identified and helped resolve any ambiguous results.

4.3.5 Study Screening

Study screening was performed in two phases. First the title and abstract of all studies identified in the systematic search were screened and discarded if they failed to meet the final eligibility criteria. Decisions on study inclusion/exclusion were made using Screening Form 1 (Appendix 4.2a). To minimise error, studies were progressed to the second screening phase if there was any uncertainty regarding their eligibility. The second screen was performed using Screening Form 2 (Appendix 4.2b) on full texts for all studies that remained following the initial screen. Each study was read in full and included in the review if it met the eligibility criteria.

4.3.6 Data Charting

Data charting is a method used in scoping reviews and is comparable to the data extraction process utilised in systematic reviews where relevant information is obtained using a common framework applied to all studies (Arksey and O'Malley, 2005). The process is iterative as pilot forms are tested on a small number of studies and are refined and retested to produce the final data charting form (Levac et al., 2010).

A single researcher (the author) conducted data charting in this review, aided by regular discussions with the larger research team. The iterative process first involved the development of a preliminary data charting form which was piloted on two studies. The number of categories

in the form was then reduced and piloted on 10 of the included studies and refined further to produce the preliminary data charting form.

Next, the preliminary data charting form was applied to all the studies and obtained detailed information on: (i) general study characteristics including the primary author, date of publication, article title, the department of the primary author and the stated purpose of study; (ii) the development literature informing the study (through a detailed written description); (iii) characteristics of the empirical model including categories for study design, the country of the study and the statistical methods used; (iv) model variables and the associated age of the participants when each variable was measured.

Finally, the data on the preliminary data charting form was condensed into a final data charting form (Appendix 4.3). Condensing the data made it easier to interpret and analyse. This was achieved by recoding the data onto a spreadsheet using Microsoft Excel. To recode the data, appropriate categories were identified by searching for common repeated responses in the preliminary data charting form. The original study data was then sorted according to these new categories. Where data could not be condensed into that category it was classified as "other". Each category included in the final data charting form is reported below alongside a brief explanation:

- *Author*: The primary author's surname.
- *Year of Study*: The year that the study was published.
- *Discipline of Study*: The department or discipline of the primary author condensed to categories of child development, economics, epidemiology, health, psychology, sociology, and other.
- *Developmental Theory:* It was not possible to clearly identify repeating categories as theory was specific and diverse. Instead development was categorised (i) according to the author's perspective on nature-nurture (categories including environmental factors only, individual factors only, or both environmental and individual factors) and (ii) according to the development domains that were included in the model (physical, cognitive, socioemotional, physical & cognitive, physical & socioemotional, cognitive and socioemotional, holistic).
- *Type of Empirical Model:* Model categories included: generalised linear models, structural equation models, survival models, dynamic factor models, growth curve models/multilevel models, and evidence synthesis models (where more than one modelling method was used).
- *Time Horizon:* The age at which model parameters were collected was recorded as the range from earliest child model input to final age of adult outcome.

- *Outcome Sector:* Each of the models' outcomes was classified as occurring in the health (yes/no), crime (yes/no), education (yes/no) or employment (yes/no) sectors.

4.3.7 Data Analysis and Results Synthesis

Data analysis mapped the general characteristics of all studies included in the review. This was performed using tables and graphically representing the condensed study information. All tables and graphs were produced on Microsoft Excel. A further analysis was conducted by cross tabulating categories form the above data charting form. This allowed patterns to be identified that may not have been apparent through the initial mapping. Cross tabulation was performed using pivot tables in Microsoft Excel. Given the large number of categories, results are not reported for cross tabulations where no relationship was identified.

4.4 Results

4.4.1 Results for the Primary Objective

As concluded in chapter two, the application in this thesis conducts economic evaluation for health centric and cross-sectoral decision perspectives, each requiring the adoption of QALYs and healthcare costs as decision endpoints. No studies were identified which specifically predicted adulthood QALYs and/or healthcare costs from measures of child development. Consequently, no studies met the full eligibility criteria for the first objective. Several studies were identified which used measures of child development to predict health and economic outcomes that are not fully relevant decision endpoints. Three of these studies have made substantial contributions to literature linking child development to adulthood outcomes and are briefly summarised below.

Firstly, Cunha and Heckman (2008) and Cunha et al. (2010) contribute to a series of studies labelled the "technology of skill formation". These models define child development as a vector of capabilities, θ_t, made up cognitive abilities (such as IQ, attention), non-cognitive abilities (such as patience, self-control, temperament and risk aversion etc.), and health stock (e.g. propensity for mortality and morbidity). The technology of skill formation incorporates central themes from the development literature, where capabilities are produced by genes, the environment and investments e.g. education, child healthcare. A multi-staged approach is used to model the child development process where a child's future vector of capabilities (θ_{t+1}) occurs as a function of their current capabilities (θ_t) and investments into these capabilities. Upon reaching adulthood θ_t is assumed to determine an individual's ability to perform tasks

and function effectively in economic and social life. The model is used by Cunha and Heckman (2008) and Cunha et al. (2010) estimate probabilities of high school graduation and adulthood earnings at age 32 in data from the USA. Findings suggest a substantial return on investments in child development particularly for educational programs delivered to children from disadvantaged backgrounds.

Secondly, the search identified the "health determinants model" by Hertzman et al. (2001) which uses linear regression models to predict adulthood health from a variety of early environmental, physical, cognitive, and socioemotional development measures in children aged 0-7. Hertzman et al. (2001) utilise data from the 1970 British Cohort Study obtaining outcome measures of self-rated health on Likert scales when individuals are aged 33. Different empirical specifications are used to account for different pathways of effect between child development and adulthood health including models which account for the latency and accumulation hypotheses (which informed the previous analysis in chapter three). The model establishes an association between adulthood health and early child development. Findings by Hertzman et al. (2001) have contributed towards a commonly cited theoretical model labelled the life-course health development framework which is advocated in Halfon and Hochstein (2002) and Halfon et al. (2014).

Thirdly, mathematical modelling is used by Layard et al. (2014) to predict how child development and early life conditions affect adulthood well-being, whilst accounting for early environmental family economic and family psychosocial factors. Layard et al. (2014) use a series of parallel regression models in the 1958 British Cohort Study to establish the pathway from measures of childhood cognitive and socioemotional development at ages 5, 10 & 16 and self-reported well-being (scored from 1-10) at age 42. The path model includes (i) direct effects between child development and adulthood well-being and (ii) indirect effects via a series of adulthood outcomes such as income, education, employment, and physical/emotional health conditions. Findings indicate that children's emotional health and early behaviour are the strong predictors of several adulthood outcomes including well-being (Layard et al., 2014).

Whilst these studies contribute significantly to linking child development to adult outcomes, they do not specifically predict adulthood QALYs and/or healthcare costs and therefore do not provide relevant measures for economic forecasting in this thesis.

4.4.2 Search and Screen Results for the Second Objective

As there were no eligible studies for the primary objective, the remainder of the results section is directed towards the second objective - a thematic analysis to identify the key modelling characteristics that should be applied when forecasting lifetime effects from intermediate

development outcomes. The database search was conducted between January 2000 and June 2016. After duplicates were removed it returned a total of 8,194 potential studies. An additional study was identified through prior knowledge of the literature. This left 8,195 studies screened for title and abstract, after which, 7,863 citations were discarded as they did not meet the eligibility criteria. Full text articles were obtained for the remaining 336 studies which were progressed to the second screen.

For the second objective, 279 studies were discarded upon reading full texts leaving 57 studies that met the inclusion criteria. The most common reasons for studies being excluded during the second screen were: no empirical model reported (40%); no relevant outcome variables at age 16 or above (34%); no relevant input variables between ages 0-12 (10%); not informed through the child development literature (4%); and any other reasons (12%). The progression of studies through the search and screening process for the second objective is illustrated in the PRISMA flow chart (Figure 4.4).

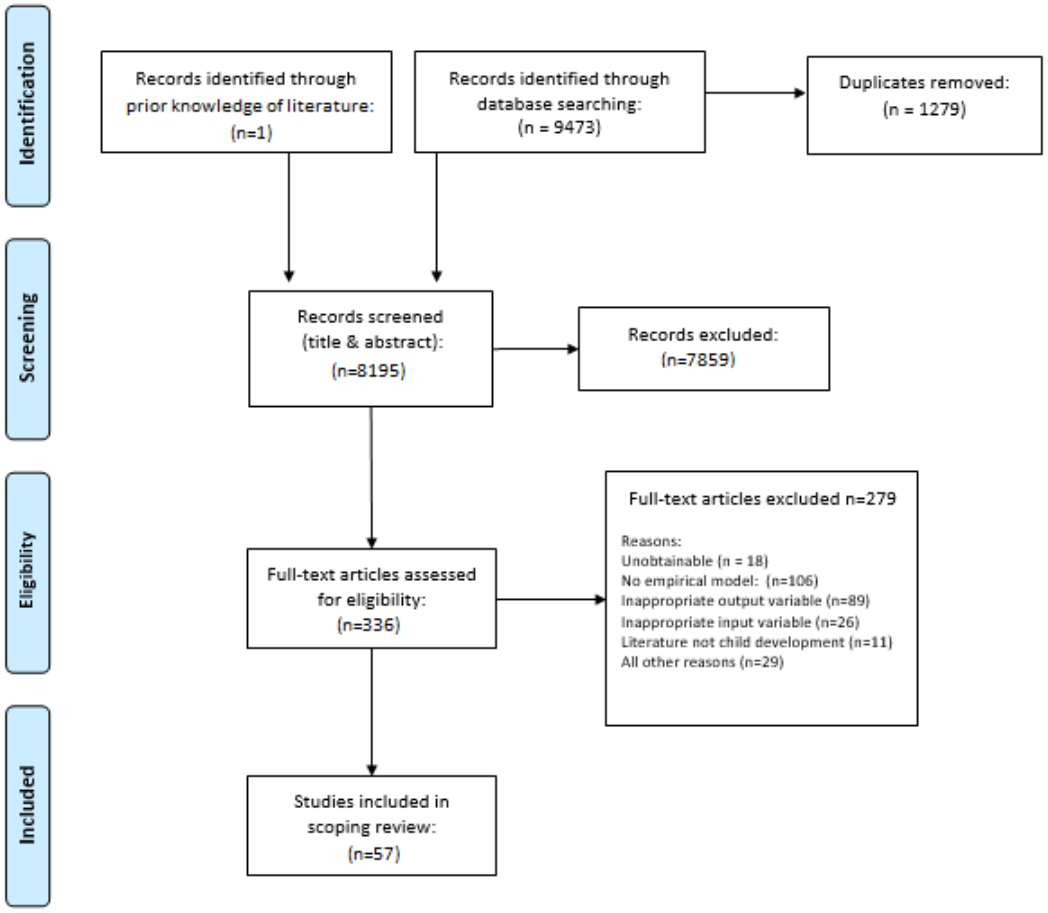


Figure 4.4*:* PRISMA flow chart displaying the results of the search and screening process for the second research objective. None of the 336 studies that passed the first screen met the eligibility criteria for the first objective.

4.4.3 Mapping Literature

4.4.3.1 General Study Characteristics

A descriptive summary of the general characteristics for each of the 57 studies included in the review is provided in Table 4.1. All studies were published between January 2000 and July 2016 and were evenly distributed between these dates. Most studies were obtained from either the USA (n=25), the UK (n=14), or Canada (n=6), and were primarily informed through the fields of psychology (n=17), economics (n=13) and epidemiology (n=8).

Table 4.1: Characteristics of Included Studies

Primary Author	Year of Publication	Location of Study	Discipline
Adair, L.	2013	Multiple locations	Health Sciences
Boyle, M.	2007	Canada	Human Development
Chandola, T.	2006	UK	Epidemiology
Chartier, M.	2010	Canada	Health Sciences
Creel, M.	2006	USA	Economics
Cunha, F.	2008	USA	Economics
Cunha, F.	2010	USA	Economics
Daniels, M.	2004	Philippines	Other
Dishion, T.	2010	USA	Psychology
Dubow, E.	2009	USA	Psychology
Engle, P.	2011	Multiple	Human Development
Fergusson, D.	2004	New Zealand	Psychology
Ferrer, E.	2004	USA	Psychology
Friedman, H.	2014	USA	Psychology
Frijters, P.	2010	UK	Economics
Hagger-Johnson, G.	2012	UK	Psychology
Hagger-Johnson, G.	2011	UK	Epidemiology
Hampson, S.	2015	USA	Psychology
Hatch, S.	2010	UK	Psychology
Healey, A.	2004	UK	Economics
Herrenkohl, T.	2010	USA	Sociology
Hertzman, C.	2001	UK	Epidemiology
Huang, C.	2011	Mexico	Public Health
Jimerson, S.	2001	USA	Psychology
Johnson, W.	2006	USA	Human Development
Kuh, D.	2004	UK	Epidemiology
Kuh, D.	2009	UK	Epidemiology
Martens, P.	2014	Canada	Public Health
Mason, P.	2010	USA	Psychology
Mason, W.	2007	USA	Economics
Moffit, T.	2011	New Zealand	Psychology
Moody-Ayers, S.	2007	USA	Epidemiology
Muennig, P.	2009	USA	Economics
Nandi, A.	2014	USA	Epidemiology
Nikulina, V.	2011	USA	Psychology
Petras, H.	2005	USA	Public Health

Table 4.1 *continued:* Characteristics of Included Studies

Primary Author	Year of Publication	Location of Study	Discipline
Reynolds, A.	2011	USA	Human Development
Risi, S.	2003	USA	Psychology
Roeser, R.	2003	USA	Psychology
Rosa Dias, P.	2009	UK	Economics
Savage, J.	2002	Multiple	Psychology
Schoon, I.	2003	UK	Psychology
Sharieff, W.	2008	Pakistan	Economics
Shen, K	2014	China	Epidemiology
Slopen, N.	2014	USA	Psychology
Te Velde, S.	2008	Netherlands	Economics
Tran, B.	2015	Canada	Public Health
Tubeuf, S.	2013	UK	Health Sciences
Vartanian, T.	2005	USA	Sociology
Wodtke, G.	2016	USA	Sociology
World Bank	2003	Multiple	Economics

All the UK studies used data from longitudinal cohorts. These included: the 1931 Lothian Birth Cohort (Hagger-Johnson et al., 2012); the 1937 UK Boyd Orr Cohort (Frijters et al., 2010); the 1946 Medical Research Council National Survey of Health and Development (Kuh et al., 2004) (Kuh et al., 2009); the Aberdeen Children of the 1950s Cohort Study (Hagger-Johnson et al., 2011); the 1958 National Child Development Study (Chandola et al., 2006) (Hatch et al., 2010), (Lindeboom et al.), (Rosa Dias, 2009); the 1958 British Cohort Study (Hertzman et al., 2001), (Tubeuf et al., 2012); the 1961 Cambridge Study in Delinquent Development (Healey et al., 2004); and the 1970 British Cohort Study (Layard et al., 2014), (Schoon et al., 2003).

4.4.3.2 Model Type

Almost all studies (n=54) used a statistical model as the selected model type. The remaining studies used an evidence synthesis framework populated with results from multiple sources of secondary data. The most common category of statistical models (n=21) were generalised linear models (GLMs) which included ordinary least squares, probit and logit models. Other model types included structural equation models (n=15), hierarchical linear models (n=5), growth curve models (n=4), dynamic factor models (n=3) and cox proportional hazard models (n=2). The statistical model used for each individual study is reported in Appendix 4.4. The distribution of model types is illustrated in Figure 4.5.

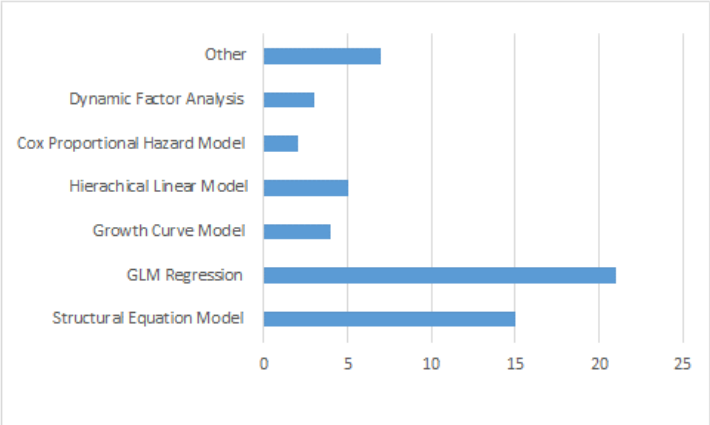


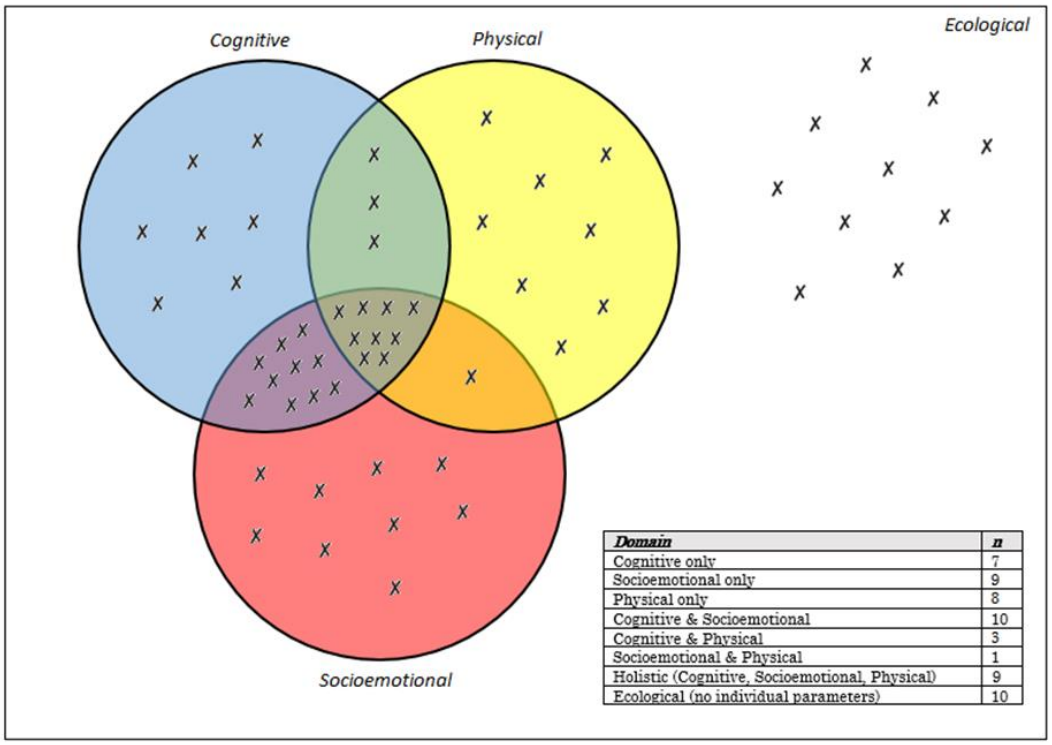
Figure 4.5: A bar chart illustrating the frequency of statistical models used across the included studies.

4.4.3.3 Development Domain

Model input parameters were defined as variables occurring in childhood (before the age of 12) and were used to predict an outcome, for example the vector of independent variables in a generalised linear regression model. Most studies (n=42) were categorised as holistic as they had input parameters that related to *both* characteristics of the child and their environment.

The commonest examples of environmental parameters were those associated with children's immediate family, where most studies captured the effects of the family economic (n=45) and/or psychosocial (n=27) environment. A smaller number of studies accounted for the wider environmental effects on child development, these including parameters relating to the children's neighbourhood (n=12) and their school/peer environment (n=6).

Each input parameter associated with characteristics of the child (rather than their environment) was classified according to whether it fell within the cognitive, socioemotional or physical development domain. Examples include IQ/cognitive performance tests (cognitive), parent/teacher reports of children's behaviour (socioemotional) and height and weight measurements (physical). In total, 29 studies reported a cognitive input parameter, 29 studies reported a socioemotional input parameter, and 21 studies reported a physical parameter. In addition, 23 studies reported parameters in more than one domain and 9 studies reported parameters in all three developmental domains. The distribution of input parameters by domain is illustrated in a Venn diagram in Figure 4.6.



Domain	n
Cognitive only	7
Socioemotional only	9
Physical only	8
Cognitive & Socioemotional	10
Cognitive & Physical	3
Socioemotional & Physical	1
Holistic (Cognitive, Socioemotional, Physical)	9
Ecological (no individual parameters)	10

Figure 4.6: Venn diagram illustrating model input parameters according to the developmental domain that they fall in. Each cross represents an individual study. Studies with parameters in multiple domains are placed in the intersection between the relevant domains.

4.4.3.4 Outcomes

The outcomes of the studies were defined as model parameters occurring in adulthood (after the age of 16) predicted or estimated from the model's input parameters, for example the dependent variable in generalised linear regression models. Outcomes were classified according to the government sector in which the outcome fell. Several studies predicted outcomes across multiple sectors using the same underlying model.

Health outcomes were most common (n=33) and included measures of self-reported health on Likert scales, presence of multiple diseases, healthcare usage, disability adjusted life year, surrogate disease markers and mortality. Educational outcomes (n=24) were typically academic test scores or total years of education achieved. Outcomes in the employment sector (n=16) related to estimations of productivity including human capital, ability to work and employment status. Crime outcomes (n=15) were usually related to incarceration rates or risk ratios for involvement in crime or risky behaviours.

4.4.3.5 Time Horizons

Model time horizons were calculated as the time from the first model input parameter (start) to the last model output parameter (end), representing the maximum time span of the model. This was available for 39 of the 57 studies. The mean start age across all studies was 6.34 years, whilst the mean end age was 33.33 years. The mean time horizon was 28.16 years, but this varied substantially as indicated by the large associated standard deviation (16.14 years) and range (62 years), Figure 4.7.

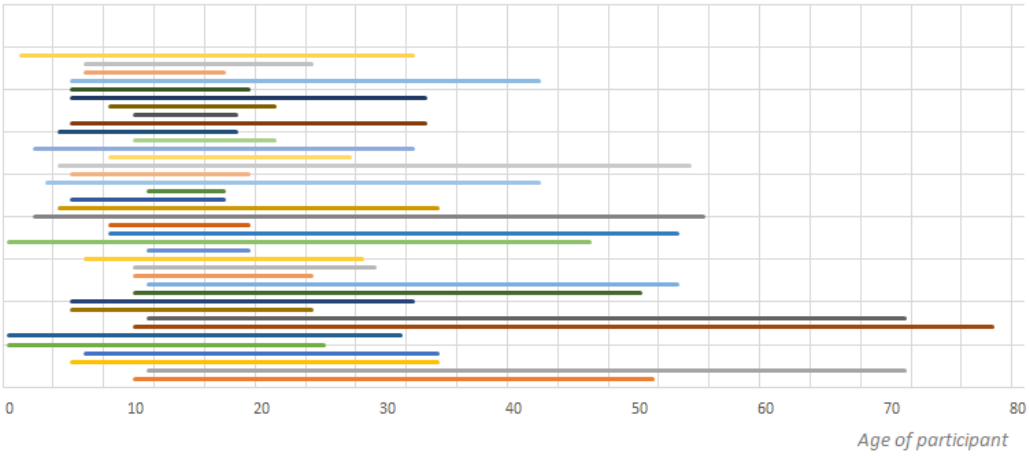


Figure 4.7: Plotted time horizon of each study. Each line represents an individual study, no colour coding scheme is applied. Bars begin at the first reported age of input and the age of final outcome.

4.4.4 Cross Tabulation Results

4.4.4.1 Development Domain and Outcome Sector

The distribution of model outcomes according to the domain of the model input parameters is illustrated in the Venn diagram in Figure 4.8. A large number (n=20) of models predicted effects on multiple outcomes occurring across different sectors. Studies were more likely to have outcomes across multiple sectors if their model had input parameters in more than one domain. This was particularly evident for the intersection between cognitive and socioemotional domains where ten of the seventeen models predicted outcomes in at least two sectors.

The outcome sector of the models appeared to be related to the development domain of their input parameters. Cognitive input parameters were estimated in nearly all models (88%) which had outcomes in the education sector, and in around half of models with crime, employment

and health outcomes. Socioemotional input parameters were frequently estimated in models with education (81%) and crime (73%) outcomes, and less frequently estimated in models with health (45%) and employment (38%) outcomes. Physical input parameters were regularly estimated in models with education (63%) and health (48%) outcomes but were irregularly estimated in models with employment outcomes (21%) and were never estimated (0%) in models with crime outcomes, Figure 4.9.

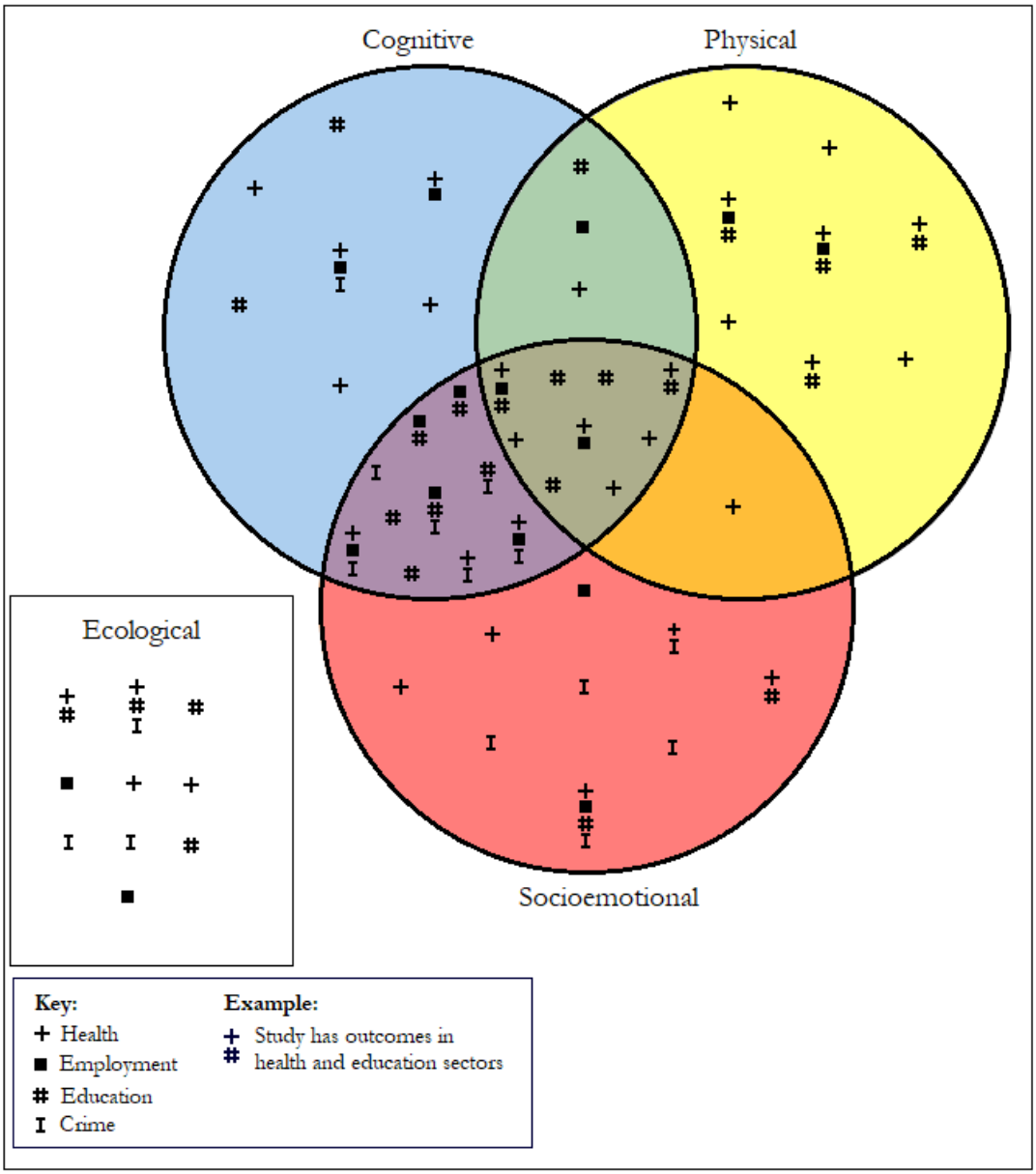


Figure 4.8: Venn diagram illustrating the development domain of the models input parameters, and the government sector of the model's outputs. Each study is represented by a combination of health, employment, education and crime outcomes as illustrated in the Key. Ecological studies are defined as those with only environmental input parameters.

The type of environmental input parameters remained fairly consistent irrespective of the model's outcome sector. For example, family economic parameters (>=80%) and family psychosocial parameters (>=48%) were frequently reported for all outcome sectors, while neighbourhood influences consistently occurred in around 20% of studies with crime, education, employment or health outcomes. There was a difference in the proportion of studies which estimated environmental parameters relating to school and peers which were more common if models had outcomes in the crime sector (20%) than in the health sector (6%), Figure 4.9.

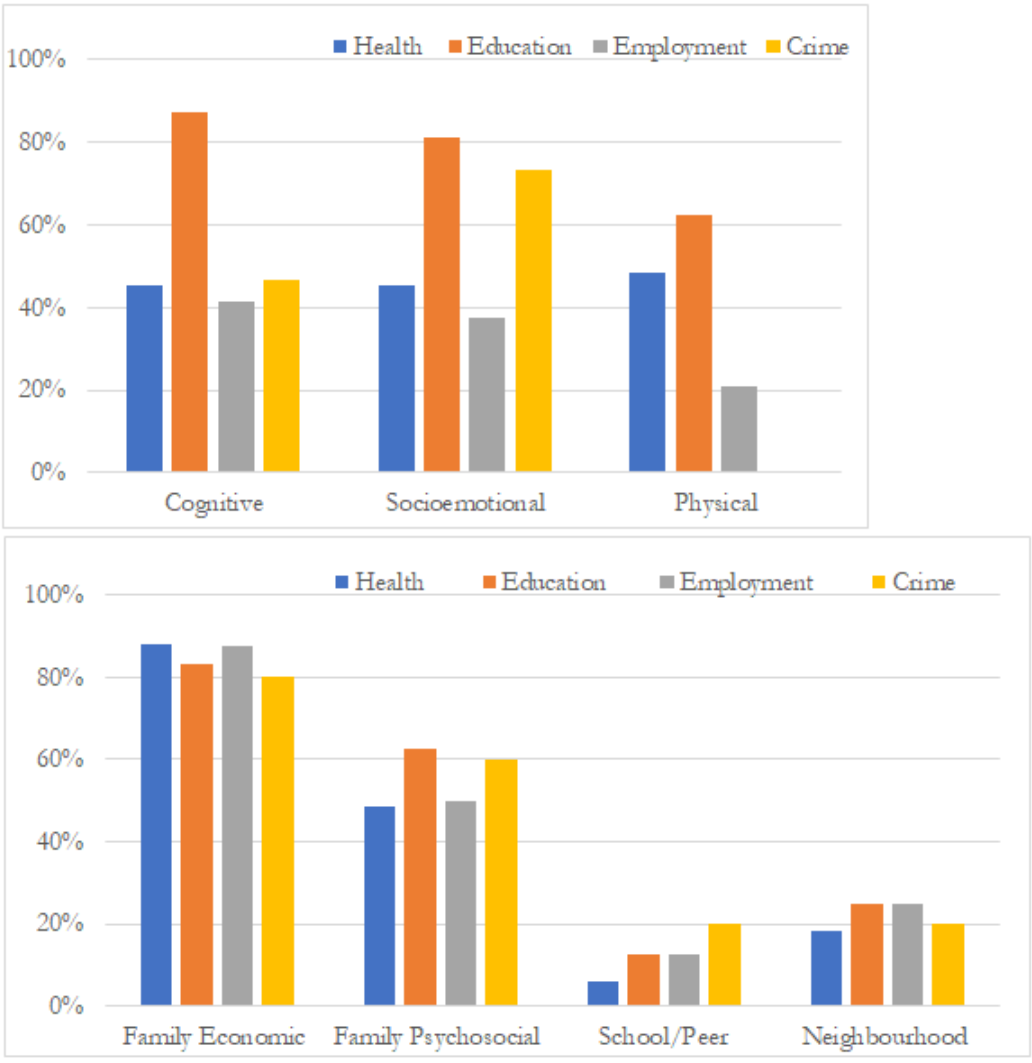


Figure 4.9: Bar charts illustrating the percentage of studies with input parameters in a specified domain, given that the model's outcome occurred in a specified sector. Percentages were calculated by dividing the total number of studies with outcomes in sector x *and* input in domain y by the total number of studies with outcomes in sector x (x and y/y). Results are obtained for individual (above) and environmental (below) input parameters.

4.4.4.2 Time horizon by development domain and outcome sector

Study time horizons varied according to the domain of their input parameters. Models with physical input parameters had the longest time horizon (mean = 32.89 years), followed by models with cognitive input factors (mean =26.69 years), while models with socioemotional input parameters had the shortest time horizon (mean = 20.70 years), Figure 4.10. Studies with cognitive and physical input parameters had time horizons that varied substantially, while less variability was present in studies with socioemotional input parameters. In addition, the mean age of model input parameters for the physical domain was younger (5.16 years) when compared with cognitive (6.19 years) and socioemotional (6.78 years) domains.

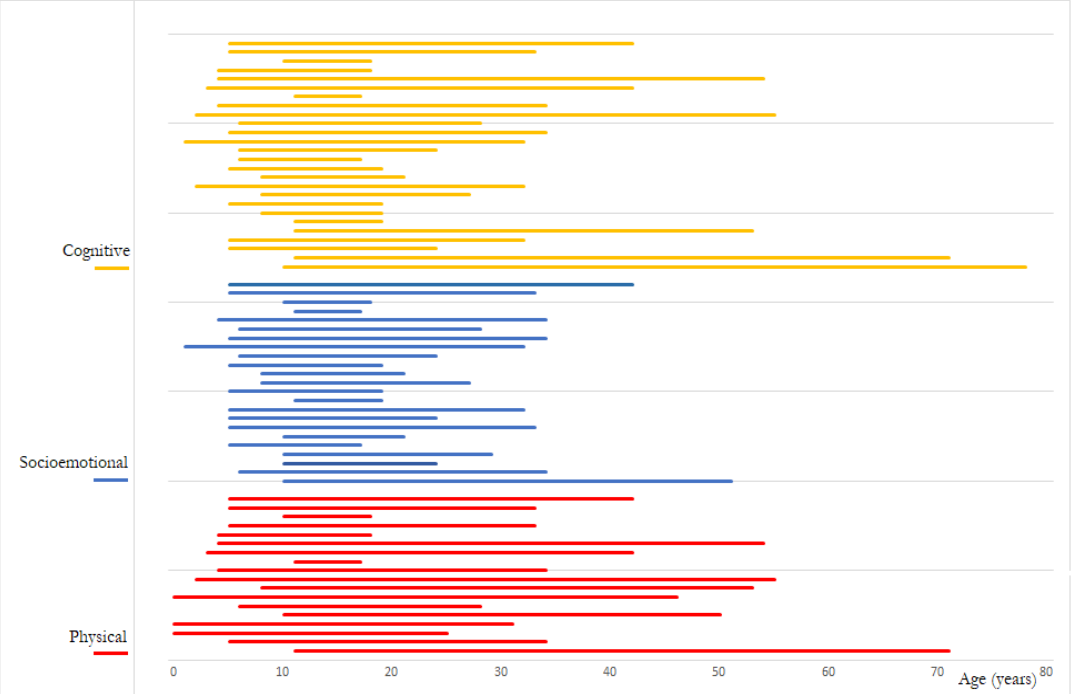


Figure 4.10: Time horizon (years) of each study by developmental domains. Each bar represents an individual study. Some studies are repeated across domains if their model input parameters occur in multiple domains.

Model time horizons also varied according to the sector in which the model's outcome fell, Figure 4.11. Models that predicted outcomes in the health sector were typically associated with a long-term time horizon with a mean equal to 36.24 years and 6 studies having final outcomes that occurred after the age of fifty. This contrasts with models that predicted outcomes in the crime sector which had the shortest mean time horizon of 17.4 years and no studies that predicted crime outcomes after the age of thirty-three. While less extreme, the time horizon of models predicting outcomes in the education sector (mean= 21.11 years) were lower than average, and the time horizon for models in the employment sector were higher than average (mean=29.5 years).

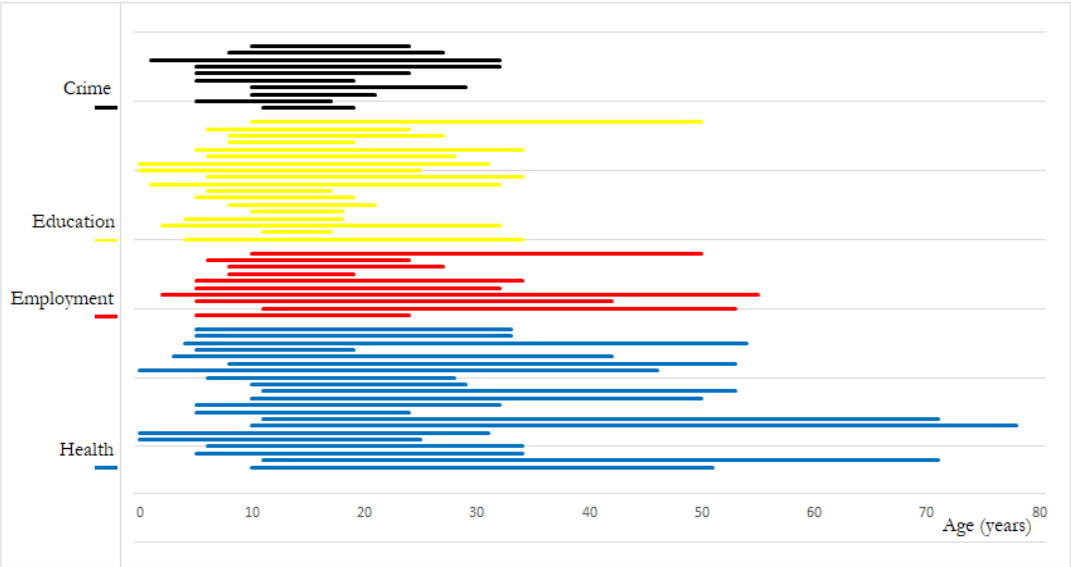
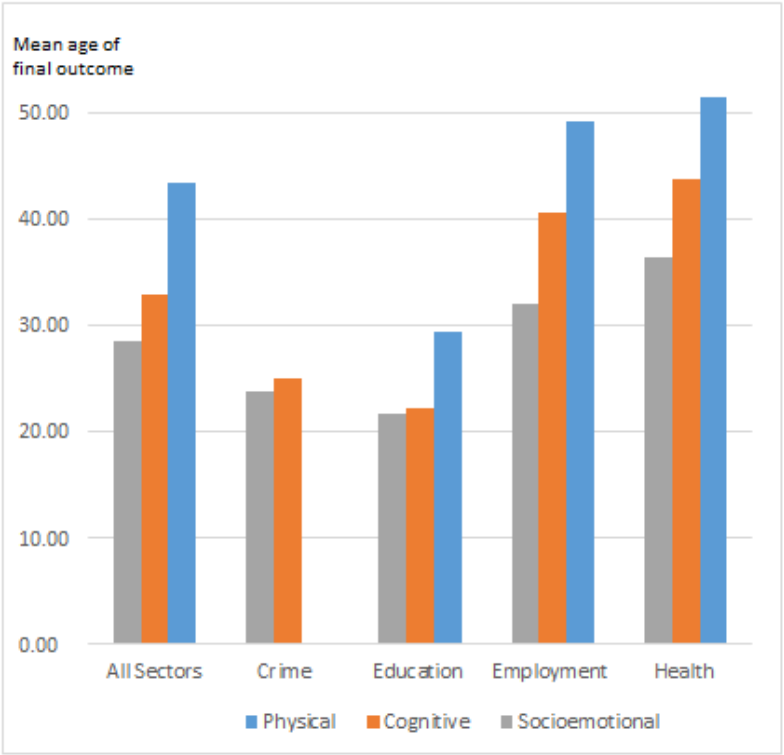


Figure 4.11: Time horizon by outcome sector. Each bar represents a single study and that some studies are repeated if they had outcomes across multiple sectors.



	All sectors	Crime	Education	Employment	Health
Cognitive	32.87	25.00	22.15	40.56	43.77
Physical	43.29	N/A	29.30	49.20	51.38
Socioemotional	28.46	23.78	21.73	36.33	36.33

Figure 4.12: Mean age (years) of final model outcomes by development domain and outcome sector.

Time horizons were further investigated by categorising the age of the final outcome according to the domain in which their input parameters fell *and* the sector in which their outcome fell. There appeared to be a multiplicative effect between domain and outcome sector. For instance, models with physical input parameters reported long-term outcomes if the outcomes occurred in the health sector (mean=51.38 years) or the employment sector (mean= 43.29 years), but not if they were in the education sector (mean=29.30). Conversely, models with socioemotional input parameters reported very short-term final outcomes in the crime sector (mean=23.46) and in the education sector (21.73) but longer-term outcomes in both health (36.33) and employment (36.33), Figure 4.12.

4.5 Discussion

4.5.1 Key Findings

This scoping review did not identify any models meeting the primary objective – being appropriate for forecasting the previously identified effects of postnatal depression on children's cognitive and socio-emotional development (chapter three) to effects on adulthood decision endpoints relevant for health economic evaluation in the UK. All studies were considered inappropriate as they did not estimate QALYs and/or healthcare costs, defined as the decision endpoints of interest in chapter two required for economic evaluation conducted from both a health and cross-sectoral decision maker's perspective.

4.5.2 Methodological Guidance for Future Research

Upon expanding the inclusion criteria, several models were identified which predicted outcomes that could be either used to partially compile QALYs (for example mortality rates might be used to determine the number of life years in the QALY calculation), or might theoretically be transformed into QALYs (for example self-rated health on an ordinal scale could, in theory, be transformed into utility scores on Visual Analogue Scales and these might be used to compile QALYs).

Whilst these studies could be useful as models for indirect estimation, a primary analysis is considered a more desirable option as: (a) partial endpoints cannot establish the full impact of child development on decision endpoints, for example, there may be stronger links between development and morbidity or healthcare usage, and with a sole focus on mortality these effects would be unobserved; and (b) transformation of outcomes to QALYs is likely to require resources for primary research which may be better spent by estimating a fully relevant model with characteristics informed through the results of this review and a complete set of decision

endpoints. Therefore, the results of the synthesis in this scoping review are used to guide the design of a primary analysis in chapter five.

As the results of chapter three identified effects of maternal depression on children's cognitive and socioemotional development, the discussion of methods is directed towards results in these development domains. The interpretations are transferable to other early life circumstances that also have effects in cognitive and socioemotional domains but may not be wholly applicable if the early life circumstances affect development in the physical domain.

4.5.2.1 Choice of Outcome Sectors

The results of this review indicate that children's cognitive and socioemotional development is likely to be associated with outcomes across multiple government sectors: models with input parameters in these domains estimated diverse effects across crime, education, employment and health. These results are consistent with the theoretical literature reviewed in chapter two which suggests that cognitive and socioemotional development is likely to have widespread effects later in life.

Economic evaluations should account for all *relevant* costs and benefits of an intervention (Drummond et al., 2015). Whether a cost or benefit is considered relevant is dependent on the perspective of the decision maker. As outlined in chapter two, this thesis conducts economic evaluation from both a health and cross-sectoral decision maker's perspective- the latter endpoints estimate QALYs healthcare costs and monetary costs/returns in other government sectors.

4.5.2.2 Model Selection

A variety of statistical methods were used by the studies in this review. The most common of these were the family of generalised linear models (GLMs). This is a class of regression models whose outcome variable (y_i) is assumed to follow an exponential family distribution, and whose mean (μ_i) is assumed to be a function of the independent variable vector (or model input parameters) ($x_i\beta$) (McCullagh, 1984).

The family of GLMs are often applicable to outcome variables with different characteristics given that different distributions can be assumed for y_i. For instance, GLMs contain the commonly used ordinary least squares, logit and Poisson/negative binomial regression models which are often an appropriate choice when modelling continuous, dichotomous and count data (McCullagh, 1984). Given their flexibility, GLMs are regularly used when estimating effects on outcomes with skewed distributions, which may include decision endpoints such as QALYs

and costs (Jones and Schoon, 2008). Accordingly, GLMs are selected as the choice of model for the analysis in chapter five.

4.5.2.3 Selection of Explanatory Variables

When estimating a GLM it is possible to both overfit and underfit the model. Overfitting is where the residual variance is represented in the model structure which can occur if an excessive number of explanatory variables are included in $x_i\beta$ which have little influence on the outcome (Burnham and Anderson, 2003). In contrast, underfitting occurs if important variables are excluded from $x_i\beta$ and might result in models failing to account for confounding factors which explain part (or all) of the relationship between the included explanatory variables and the model's outcome (Chen et al., 1999). It is often most desirable to estimate a parsimonious model which achieves the required level of explanation with as few (explanatory) variables as is needed (Vandekerckhove et al., 2015).

The results of this review can be used to identify/exclude the important/irrelevant explanatory variables to estimate a parsimonious model in chapter five. In general, the results are consistent with the theoretical literature reviewed in chapter two suggesting that development is holistic occurring across multiple domains (Overton, 2010b), (Santrock, 2003) and being influenced by a wide range of environmental factors (Bronfenbrenner, 1992). Variables for (i) cognitive & socio-emotional development and (ii) family economic, family psychosocial and neighbourhood environmental factors were consistently estimated across all outcome sectors and should be included in future models estimating outcomes in crime, education & employment and health.

In contrast, physical development was only found to be a typical input parameter for models with health outcomes and not included as an explanatory variable for future models estimating outcomes in crime and education & employment. Meanwhile, school/peer environmental factors were infrequently estimated for education, employment and health sectors, and are only used as explanatory variables in chapter five for models predicting crime outcomes.

4.5.2.4 Study Sample

Economic evaluations should be relevant to the perspective of decision maker (Urdahl et al., 2006). This means that evidence to inform economic evaluation should have external validity and be generalisable to the target population of the health policy (Carlson and Morrison, 2009). As this thesis is aimed at informing UK policy making, a relevant study population ought to be estimated in a sample that is geographically representative of the UK population. The results section of this review summarised all UK samples used within the included studies, and these were considered as potential data sources for the analysis in chapter five.

Evidence obtained from longitudinal observational studies might lack external validity if the study is conducted using historical data occurring a long time before the date of the policy decision as the demographics of the target population might change between these dates (Carlson and Morrison, 2009). As indicated in the results of this review, this is particularly relevant for models forecasting lifetime outcomes from childhood input parameters, which typically span multiple decades.

Therefore, the most appropriate study sample for the analysis in chapter five was selected by identifying a UK based study which is conducted closest to the date of the decision (2018) whilst also containing an appropriate time horizon – this was assumed to require adulthood outcomes collected in the fourth decade as this was the longest mean final age of outcome identified in the review (equalling 43.77 for health outcomes). The British Cohort Study 1970 was selected as most appropriate for the analysis in chapter five as the latest sweep was conducted in 2012 when participants were aged 42.

4.5.3 Strengths and Limitations

4.5.3.1 Strengths

This scoping review benefited from the application of rigorous methodology throughout which followed published guidelines set out by Arksey and O'Malley (2005), Colquhoun et al. (2014) and Levac et al. (2010). This included: the design of a comprehensive search strategy and eligibility criteria that was aided by discussion with a qualified librarian; screening studies using predesigned screening forms; charting data across predesigned and pretested data charting forms; and systematically analysing charted data using a common analytic strategy.

The review also captured an appropriate range of the development literature as the search strategy was conducted across six diverse electronic databases. This is exemplified by the inclusion of studies that are informed through multiple disciplines such as child development, economics, epidemiology, health sciences and psychology.

It was possible to map a large volume of the literature by streamlining the data charting process which extracted only the key study information rather than obtaining a comprehensive list that described each study in detail. Limiting the study information and analysis to a smaller number of categories enabled key themes to be identified that were useful when informing future research objectives.

4.5.3.2 Limitations

This scoping review was limited by resource constraints which meant that the review was conducted by a single researcher. Scoping reviews should ideally be conducted by a research team where multiple individuals are involved in study screening, study selection, and data charting processes to avoid errors and enable discussion of ambiguities (Arksey and O'Malley, 2005). It is possible that errors could have led to relevant studies being excluded from the review, which may have affected the conclusions drawn for both research objectives. Therefore, uncertainties were discussed throughout with a larger research team to minimise the potential occurrence of error.

In addition, as research was only conducted by one individual, the review had time constraints which resulted in literature searching being limited to electronic databases and to studies written in the English language. No grey literature was searched, and no other means of searching was applied e.g. hand searching journals or searching reference lists. It is possible that supplementary searching methods would have identified additional studies that met the inclusion criteria for this review.

The review may also be limited by the eligibility criteria which excluded models of single diseases/health conditions. This was done to increase the specificity of the search strategy. As stated by Pham et al. (2014) scoping reviews provide a general account and are not intended to be an exhaustive list of the literature. The exclusion of some studies that may have been identified through less restrictive eligibility criteria is unlikely to have affected the results of the synthesis.

Finally, this scoping review did not attempt to provide any form of critical appraisal. However, Arksey and O'Malley (2005) suggest that the role of a scoping review is not to assess study quality but should instead focus resources on mapping and describing the key characteristics of the literature. As the studies in this review were used to inform study design rather than being used directly as sources of evidence, the impact of not providing critical appraisal is likely to be minimal.

4.5.4 Conclusions

This chapter described a scoping review aiming to identify a mathematical model to be used to forecast the effects of maternal depression on lifetime decision endpoints, based on children's cognitive and socioemotional development outcomes measured in chapter three. The search strategy was conducted across six electronic databases and identified 8,195 studies, none of which were considered appropriate for the primary research objective. Upon relaxing the

eligibility criteria, 57 studies were found to estimate clinical health or economically relevant endpoints. These studies were analysed to identify the key modelling characteristics that should be applied when forecasting lifetime effects from intermediate development outcomes. The results informed the study design of the analysis presented in chapter five.

Chapter 5: A Mathematical Model Predicting Lifetime Outcomes from Measures of Child Development

5.1 Summary

The overarching aims of this thesis focus on development of a methodology which estimates the lifetime effects associated with early life circumstances, and determination of whether lifetime estimates influence the cost-effectiveness results in an applied health economic evaluation. Chapter four began to address the second stage in the estimation of lifetime effects, suggesting mathematical modelling as an appropriate method to forecast lifetime decision endpoints from measures of child development, whilst also identifying key modelling criteria. This chapter builds on these results by estimating a set of generalised linear regression models which establish the relationship between children's cognitive and socioemotional development at age ten and lifetime decision endpoints relevant for economic evaluation in the UK. Combining the results from the applied example in chapter three with the predictions from these models, this chapter conducts the full indirect estimation methodology, forecasting the effects of postnatal depression on children's lifetime Quality Adjusted Life Years, healthcare costs, costs/returns to the education sector, and costs/returns to criminal justice sector. The applicability of this model to other research settings (i.e. forecasting effects for other early life circumstances) is discussed. Lifetime estimates from this chapter are used directly as parameters in a decision model which assesses the cost-effectiveness of screening for postnatal depression in chapter six. Figure 5.1 illustrates the contributing role of evidence from this chapter in the overall methodology estimating lifetime effects in this thesis.

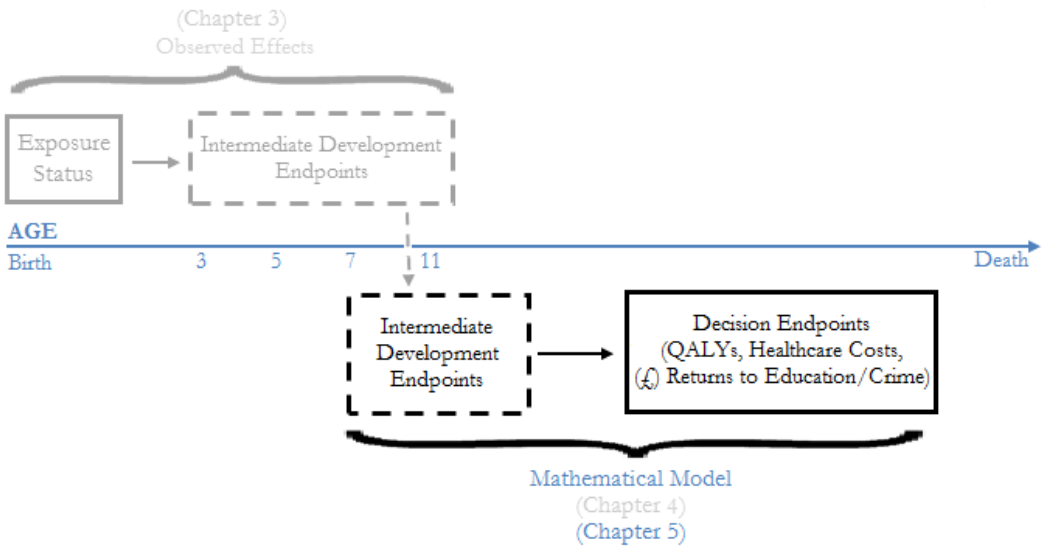


Figure 5.1: Demonstrates how this chapter contributes to the overall estimation of lifetime effects.

5.2 Introduction

5.2.1 Structure of a Mathematical Model for Indirect Estimation

Chapter four proposed the second stage of indirect estimation could be most appropriately conducted using a mathematical model to establish the relationship between intermediate measures of child development and lifetime decision endpoints. As demonstrated in Figure 5.2 the intermediate cognitive and socio-emotional development outcomes enter the mathematical model as input parameters. The input parameters are then transformed to the relevant lifetime endpoints through a series of modelling equations.

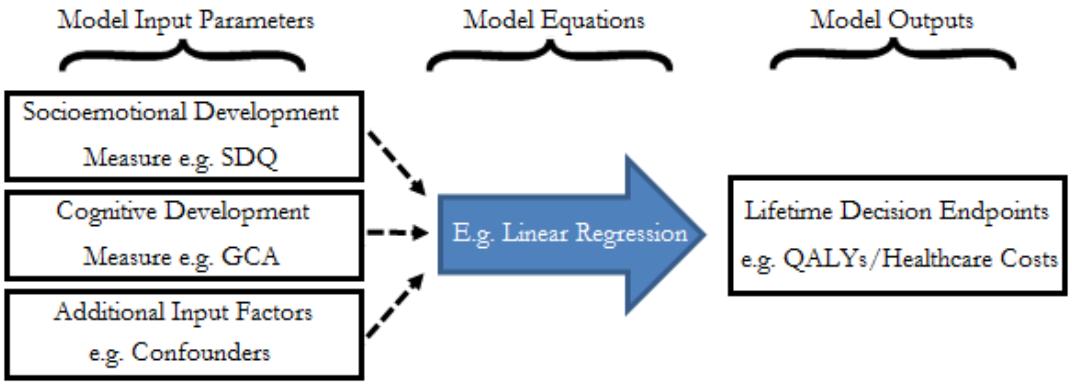


Figure 5.2: Illustrates the structure of a mathematical model used to forecast lifetime effects for indirect estimation. Cognitive and socioemotional development measures are entered as (two of several) input factors into modelling equations which output lifetime decision endpoints.

Results from chapter four suggest that the category of generalised linear models (GLMs) could provide an appropriate type of model to forecast lifetime effects during indirect estimation. Lifetime estimates could be obtained using this category of model by:

(i) Identifying the incremental treatment effects of postnatal depression on both cognitive (β_1) and socioemotional (β_2) intermediate outcomes (chapter three results);

(ii) Estimating a GLM with cognitive (X_1) and socioemotional (X_2) development in the explanatory variable vector;

(iii) Obtaining lifetime predictions for the incremental effects of postnatal depression by multiplying the coefficients attached to X_1 and X_2 by β_1 & β_2.

This approach is advocated by Layard et al. (2014) when predicting the effects of early life health policies on lifetime well-being endpoints.

For indirect estimation to be appropriate the same intermediate constructs are required. That is, the development measures adopted when identifying incremental effects (chapter three) should align with the development measures entered in the forecasting model as input parameters (or explanatory variables). For example, there would be little sense in estimating effects of postnatal depression on children's vocabulary at age 11 and using a different measure, say numerical ability at age 5, to predict lifetime outcomes. To remain consistent with previous research in chapter three, an appropriate model requires measures for cognitive development to be specified as General Cognitive Ability (GCA) informed through g-factor and CHC theory (Hill, 2005), and socioemotional development to be measured on the SDQ or a similar maternally reported and empirically validated child behavioural screening questionnaire.

5.2.2 Model Outcomes for Economic Evaluation

If the purpose of a mathematical model is to forecast outcomes to inform economic evaluation, then model outputs should be tailored to the requirements of the decision-maker. Ideally, economic evaluation identifies the effects of health technologies on *all* the outcomes relevant and influential to policy decisions (Drummond et al., 2015). As was identified in chapter two, the relevant decision endpoints for economic evaluation in the UK according to a health centric decision maker's perspective include Quality Adjusted Life Years (QALYs) and healthcare costs. If adopting a cross-sectoral decision maker's perspective, decision endpoints should also include additional costs/cost savings in sectors outside of health and a measure of economic productivity. According to the results of chapter four, the adulthood effects of child development may be expected to occur in crime, education, employment and health sectors.

The identification of *all* relevant and influential outcomes also requires costs/benefits to be identified over the appropriate time horizon (Drummond et al., 2015). The literature review in chapter two and the results of chapter four suggest that the costs and benefits of childhood health technologies are likely to continue throughout the lifespan, particularly for employment and health outcomes. The most appropriate decision endpoints for the evaluation of early childhood health policy ought, therefore, to be outcomes measured across the entire lifespan. However, due to excessive costs, it not usually possible to conduct studies which follow up participants until their death. To account for truncated study time horizons, economic evaluation can use methods to extrapolate the decision endpoints observed within studies to endpoints across the remainder of their lifetime (Garnett et al., 2011). One example is the use of survival analysis to extrapolate from observed to lifetime mortality rates (Jackson et al., 2016).

Regardless of outcome type, it is generally agreed that decision endpoints should be discounted if they are expected to occur over several years after the implementation of the health technology. Discounting might be applied to future decision endpoints for a variety of reasons including opportunity costs associated with investment now (i.e. positive rates of return); a time preference for consumption sooner rather than later; catastrophic risk; and diminishing marginal utility (Attema et al., 2018), (Severens and Milne, 2004). Policy makers may not agree on the level of discount rates or whether the same rates ought to be applied to both costs and benefits (Attema et al., 2018), (Drummond et al., 2015). To account for this uncertainty economic evaluations are often reported in scenario analyses which apply different discount rates. For example, the NICE (2013) reference case suggests cost-effectiveness results ought to be reported with costs and benefits discounted at both 3.5% and 1.5%.

5.2.3 Research Objectives

Within the context of the overall thesis aim, to estimate a mathematical model which can be used to forecast lifetime effects from measures of child development, the specific objectives of the chapter are to:

1. Predict the effects of a unit change in measures of cognitive and socioemotional development on the following adulthood decision endpoints:
    a. QALYs
    b. Healthcare costs
    c. Monetised costs/returns to the criminal justice sector
    d. Monetised costs/returns to the education sector[6]
2. To ensure the results of the study can be used to inform economic evaluation in a variety of settings, a further objective of the research is to identify predictions for model outputs which differed in terms of:

(i) Time horizon (observed outcomes vs. extrapolated lifetime outcomes)

(ii) Discount rate (0%, 1.5%, and 3.5%)

Upon achieving these objectives, this research is combined with results from chapter three, enabling the indirect estimation of lifetime decision endpoints associated with children exposed to symptoms of postnatal depression. The research in this chapter adds to a sparse scientific evidence base: chapter four did not identify any mathematical models which predicted lifetime QALYs and healthcare costs from measures of child development.

[6] Monetised costs and returns in education include effects to economic productivity. This is further expanded upon in the research methods (section 5.3).

5.3 Methods

5.3.1 Study Sample

This analysis uses data from the 1970 British Cohort Study (BCS 1970), a longitudinal study which obtained information on over 17,000 participants born in England, Scotland and Wales. The choice of study sample was informed through the results of chapter four, which suggested the BCS 1970 as the longitudinal study most likely to provide temporally relevant data for UK health policy over an appropriate time horizon. The BCS recruited participants in a birth survey during 1970 and administered follow up sweeps when they were aged: 5, 10, 16, 26, 30, 34, 38 and 42. The childhood surveys utilised parental questionnaires obtaining information relating to children's health, their physical, cognitive and socioemotional development, and the social and economic environment surrounding the child. The surveys conducted in adulthood consisted of detailed questionnaires which asked participants to self-report their employment, education, health and social circumstances (Elliott and Shepherd, 2006).

5.3.2 Model Input Variables

5.3.2.1 Socio-emotional Development

Socio-emotional development was measured in the BCS 1970 on a shortened version of the Rutter Parental "A" Scale of behaviour (RBS) when children were aged 10. The full RBS is a parental questionnaire containing 31 items which relate to eight childhood problems (e.g. headaches and bedwetting), five difficulties (e.g. speech and eating disorders) and eighteen undesirable behaviours (e.g. aggression, bullying and social isolation). Parents are asked to score each item in terms of the frequency at which the child exhibits the described behaviour, possible responses (and scores) include: "never applies" (=0), "sometimes applies" (=1), and "certainly applies" (=2). The shortened RBS in the BCS 1970 contained 19 items, which were summed to produce a continuous measure ranging from 0-38. A full description of the items included in the shortened RBS are provided in Appendix 5.1.

Primary results for this chapter are reported for a standardized RBS variable (mean=0, standard deviation=1). In addition, when predicting lifetime effects, the RBS variable was transformed to the same scale as was used for SDQ total difficulties in chapter three. Transformation was achieved by multiplying the standardised RBS scores by a value of forty (the range of the scale for SDQ scores). This ensured indirect estimation was conducted by linking studies with consistent measures of child development i.e. measures of the same socioemotional construct obtained on the same scale.

5.3.2.2 Cognitive Development

This research measured cognitive development as General Cognitive Ability (GCA) when children were aged 10. Remaining consistent with methodologies from chapter three, it was assumed that GCA could be identified in the BCS 1970 by conducting principal component analysis (PCA).[7] Parsons (2014) has previously used PCA to identify a single factor of intelligence that explained 58% of the correlation between test scores in the BCS 1970. PCA was informed by eight psychometric tests which assessed a variety of narrowed cognitive abilities in the BCS 1970 when children were aged 10. Each test was self-completed by children and included: the Edinburgh Reading Test; the Friendly Maths Test; The Pictorial Language Comprehension Test; the Spelling Dictation Task; and four British Ability Scales subtests for Word Definitions, Word Similarities, Recall of Digits, and Matrices. Following PCA, the derived GCA variable was transformed to the same scale as is commonly used to represent IQ, and which was adopted for the growth curve analysis in Chapter 3 (mean=100 and standard deviation=15). Descriptions of all the psychometric tests administered in the BCS 1970 are summarized in Table 5.1 and described in detail by Parsons (2014).

Table 5.1: Description of the Variables Used for PCA to Identify GCA in the 1970 BCS

Variable	Cognitive Abilities Assessed
Shortened Edinburgh Reading Test	Word recognition: vocabulary, syntax, sequencing, comprehension and retention.
Friendly Maths Test	Arithmetic, number skills, fractions, algebra, geometry and statistics.
Pictorial Language Comprehension Test	Spelling and phonetic decoding.
Spelling Dictation Task	Vocabulary, sentence comprehension, and sequence comprehension.
British Ability Scales Word Definitions	Verbal: Understanding of word definitions.
British Ability Scales Word Similarities	Verbal: Identifying patterns between words.
British Ability Scales Recall of Digits	Non-verbal: Memory and attention.
British Ability Scales Matrices	Non-verbal: Pattern and shape recognition.

Notes: All variables obtained from the 1980 sweep for children aged 10

[7] Refer to Chapter 3 section 3.3.2.2 for a more detailed discussion regarding the use of Principal Component Analysis to identify General Cognitive Ability

5.3.2.3 Covariates

The models' covariates were identified from questionnaires delivered to BC1970 study participants and their parents during the first three sweeps, when cohort members were aged 0- 10. The covariates related to individual and environmental factors which could potentially influence adulthood outcomes through their effects on the lifespan development process. The selection of covariates was informed by the results of chapter four, and specifically a study by Layard et al. (2014) who used individual, family economic and family psychosocial covariates in a model which estimates adulthood well-being from measures of child development in the BCS 1970.

Individual covariates included dummy variables obtained during the birth sweep for biological sex (female/male) and whether cohort members were ethnically white British (yes/no). The remaining three individual covariates related to the participant's health status at age 10: a Body Mass Index variable was derived by dividing measures for weight (kg) by the square of height (m); a dummy variable for hypertension (yes/no) was identified according to whether systolic blood pressure measurements at age 10 exceeded specific hypertension thresholds for height and sex (as reported in Falkner et al. (2004)); and a dummy variable for poor health status (yes/no) was defined according to whether participants were above the 75th percentile (>5) for total number of childhood disease according to self-reports of: asthma, bronchitis, eczema, hay fever, hearing loss, inguinal hernia, middle ear infection, pathological heart condition, pneumonia, recurrent abdominal pain, sore throat, and urinary tract infection.

Environmental covariates included family economic, family psychosocial & neighbourhood conditions in addition to factors relating to each participant's peer group. Family economic conditions were approximated based on parental income and educational status which are common measures used in child development research (Bradley and Corwyn, 2002). This analysis identified family income as the natural logarithm of both parents self-reported gross earnings. Education status was established using a continuous variable for the total number of years each participant's father had been in education.

Four variables were included to capture the effects of the *family psychosocial* environment. The first was a dummy variable for a stable family environment (yes/no) defined if the participant's parents were married during the birth sweep and remained married during the age 10 sweep. Second, a continuous variable established the total number of siblings a child had at birth. Thirdly the level of overcrowding in a participant's household was estimated at age 10 as a continuous variable, by dividing the number of people living in a participant's household by the total number of household rooms. The final psychosocial variable was a continuous covariate identifying the frequency of participant engagement in family activities when they were aged 10.

The BCS 1970 parental questionnaire asked mothers to report (yes/no) whether family activities included walking, outings, eating together, holidays, shopping, talking for at least five minutes, or eating out at restaurants. The frequency of family activities was derived by summing the total number of positive responses across all items resulting in continuous variable which ranged from 0-6.

The quality of the early *neighbourhood* was identified through questionnaires administered to BCS 1970 interviewers during the age 5 sweep when they visited the cohort member's household to deliver the parental questionnaire. This analysis specified a categorical variable for neighbourhood quality by ranking the interviewers response categories: Poor quality neighbourhoods included those stated as "poor quality or closely packed" houses; average quality neighbourhoods were those which contained "council houses or flats" or "market or rural communities"; whilst good quality neighbourhoods were those which were categorised as "well to do, well-spaced and well maintained houses".

The final covariate assigned participants as belonging to *risky peer groups* using a dummy variable (yes/no) for peer group smoking status. This was identified at the age 10 sweep where participants were directly asked "how many of your friends smoke cigarettes?" – Answers of "most of them" and "some of them" were assumed to positively indicate peer smoking.

As was alluded to in the results from chapter four, the impact of individual health, family economic, family psychosocial, neighbourhood and peer covariates may affect each of the crime, employment, and health outcomes differently. For example, the results of the scoping review suggested neighbourhood and peer influences to be particularly important in predicting future crime participation, while childhood health status was less influential. To ensure models were parsimonious, covariates were limited to those that were identified as important predictors in chapter four. Table 5.2 summarises the covariates used for each outcome sector.

Table 5.2: Summary of Model Covariates According to Outcomes Sector

	Health	Education & Employment	Crime
Gender & Ethnicity	Yes	Yes	Yes
Individual Health	Yes	No	No
Family Economic	Yes	Yes	Yes
Family Psychosocial	Yes	Yes	Yes
Neighbourhood	Yes	Yes	Yes
Peer Smoking	No	No	Yes

5.3.3 Health Outcome Variables

5.3.3.1 General Approach to Identify Health Outcomes

The research objectives required the identification of two health outcomes: quality adjusted life years (QALYs) and healthcare costs incurred by the NHS and Personal Social Services (PSS). Typically, QALYs and healthcare costs are obtained through direct responses from participants (Drummond et al., 2015). For instance, QALYs combine length of life with a measure of health-related quality of life (HRQoL) which is most commonly obtained from preference-based questionnaires such as the EuroQol-5D (EQ-5D), the short form 6D (SF-6D), and the health utilities index (HUI). Meanwhile, healthcare costs might be best obtained by administering healthcare resource usage questionnaires and multiplying each healthcare item used by an appropriate price for that resource (Drummond et al., 2015), (NICE, 2013).

The questionnaire items typically required to identify QALYs/healthcare costs were not collected in the BCS 1970. Therefore, health outcomes were approximated by assigning mean HRQoL decrements, mean life years, and mean healthcare costs to each participant who reported suffering from any of the following five chronic health conditions: cancer, depression, diabetes, hypertension and obesity. The five diseases were selected as each has plausible links with child development being causally associated with lifestyle factors (Doll and Hill, 1956), (DPPRG, 2002), (Krotkiewski et al., 1979) which have antecedents in early childhood (Lerner and Busch-Rossnagel, 2013), (Overton, 2003). Further, the five diseases were assumed to provide an indication of general health in the population as each is associated with an extremely high health and economic burden (Newton et al., 2015), (Wang et al., 2011).

The presence or absence of cancer, depression diabetes and hypertension were identified based on self-reports (yes/no) in health status questionnaires. Meanwhile, obesity was assigned if participants BMI was greater than 30, with BMI being derived directly from height and weight measures during the BCS 1970 follow-up interviews. All health outcomes were established for participants at ages 26, 30, 34, 38 and 42.

5.3.3.2 Health Related Quality of Life

Mean HRQoL was obtained in a study by Sullivan et al. (2011) who report a catalogue of EQ-5D scores based on UK preferences for 135 chronic (three digit) International Classification of Disease (ICD)-9 codes from 79,522 participants in the US based Medical Expenditure Panel Survey. Participants without disease were assumed to achieve a "well" HRQoL equal to 0.83 which corresponds with the mean HRQoL across all individuals in Sullivan et al. (2011). Participants with disease were assigned a HRQoL by subtracting the

HRQoL decrements associated with disease from the "well" HRQoL score. The HRQoL for well and diseased participants is reported in Table 5.3.

Table 5.3: Mean Health Related Quality of Life per ICD-9 Code

Disease	ICD-9 Code	Mean HRQoL Decrement for Disease	Mean HRQoL
No Disease	N/A	N/A	0.830
Cancer	140-239*	0.04179	0.788
Depression	296	0.12691	0.703
Diabetes (type 1 and type 2)	250	0.05319	0.777
Hypertension	401	0.04604	0.784
Obesity	278	0.07086	0.759

Notes: All data is sourced from Sullivan et al. (2011)
*mean HRQoL taken across all ICD codes for Neoplasms

As the BCS 1970 did not provide full information on participants' cancer type it was considered most appropriate to identify a weighted mean utility across all the ICD-9 codes for cancer (all neoplasms ICD-9 140-239). The grouping of benign and malignant cancers, along with cancers in and out of remission resulted in relatively low overall reduction in health utility (0.042). Similarly, the BCS 1970 study does not distinguish between type one or type two diabetes, therefore, a grouped HRQoL decrement was obtained for diabetes using ICD-9 code 250. Specific utilities were available for both obesity (278) and hypertension (250). Meanwhile, the most appropriate three-digit ICD-9 code for depression was the grouped code for all affective mood disorders (296) – this corresponds with the more specific ICD-10 codes for single episode (F32) and recurrent major depressive disorder (F33) (Fiest et al., 2014).

Comorbidities were accounted for using additive HRQoL decrements if participants suffered from more than one of the five diseases; the study by Sullivan et al. (2011) used multivariate regression methods which adjusted utility values for comorbidity meaning HRQoL decrements can be added without risk of double counting. No multiplicative effects for comorbidities were included. Health utility scores were applied equivalently regardless of the age at which disease occurred as the multivariate regression by Sullivan et al. (2011) did not find a significant or substantial relationship between age and HRQoL.

5.3.3.3 Expected Life Years

Life expectancies were estimated based on gender and participants' disease status at the final sweep (age 42). Non-diseased participants were assumed to achieve the gender specific mean UK life expectancy for individuals born in 1970 (male=81 years, female =84 years) (*ONS*, 2017). Life expectancies for participants reporting a disease at age 42 were calculated using evidence from a large population survey by the NHS Executive (1996) which utilised UK data from multiple secondary sources including Hospital Episode Statistics, OPCS Morbidity

Statistics in General Practice, and Death Registration Data. The report obtained information on age and cause of mortality to calculate the percentage of life years lost for specific ICD-9 codes.

Life years lost statistics establish the average years an individual would have lived if they had not died prematurely (Martinez et al., 2019). Specifically, the NHS Executive (1996) report calculated: *total life years lost per individual* by subtracting their age of death from a reference age representing a normal life expectancy (assumed equal to 75); *percentage life years lost per individual* by dividing their total life years lost by the reference age; and *percentage life years lost per disease* by calculating the mean of percentage life years lost across all individuals with cause of death assigned to the relevant ICD-9 code. The NHS Executive (1996) data was selected as it was the only study which consistently estimated life years lost for all five diseases included in this analysis.

Final life expectancies for each individual reporting disease at age 42 are described in Table 5.4. These were calculated by multiplying the % life years lost per disease ICD-9 code from the NHS Executive (1996) report with average ONS (2017) population life expectancies estimates by gender (male=81 years, female =84 years). Equivalent ICD-9 codes/groups were used to classify diseases as were used for HRQoL. Comorbid participants were assigned the lowest life expectancy for the diseases they reported.

Table 5.4: Mean Life Expectancy by Disease and Gender

Disease	ICD-9 Codes	% Life Years Lost[1]	Mean Life Years (Males)	Mean Life Years (Females)
No Disease	N/A	N/A	81.00	84.00
Cancer	140-239*	2.18	79.23	82.17
Depression	296	0.04	80.97	83.97
Diabetes (type 1 and type 2)	250	0.92	80.25	83.23
Hypertension	401	0.44	80.64	83.63
Obesity	278	0.63	80.49	83.47

Notes: (1) Values obtained from NHS Executive (1996). The % years of life lost represents the expected mean reduction in length of life for individuals with disease when compared to non-diseased individuals. Values obtained for population aged 75.
*Mean obtained across all ICD codes for Neoplasms.

Ideally, life expectancies would have been calculated using data on % life years lost in individuals with a disease onset corresponding with the age of members in the 1970 British Cohort (i.e. onset < 42). This information was not available in the NHS Executive (1996) data where life years lost were estimated for all individuals aged 0 – 75. The implication of this limitation is discussed in Section 5.5.4.

5.3.3.4 Quality Adjusted Life Years

Within study QALYs were calculated by aggregating HRQoL utility weights for each participant across the study period. Aggregation was achieved using the trapezium rule (Equation 5.1) to estimate the integral, or area under the curve, over the 16 years during which HRQoL data was available (Brazier et al., 2017). Including potential comorbidities there were 30 possible HRQoL values per sweep, four follow up sweeps, and once aggregated a maximum of 810,000 (30^4) unique QALYs values. Given the large number of potential values the QALYs variable was assumed to be continuous.

Equation 5.1: The Trapezium Rule to Estimate the Area under a Curve

$$\int_{x_0}^{x_n} f(x)dx = 1/2h[(y_0 + y_n) + 2(y_1 + y_2 + \cdots y_{n-1})]$$

Where:

y = the value of the function f(x) at time interval x_0, x_1, …, x_n

h = constant equal to duration of time between intervals x_0 & x_1, x_2 & x_3, …, x_{n-1} & x_n

A small number of participants died during the study, and this ought to have been captured in the QALYs outcome. The identification of deceased study members requires a special licence which could not be obtained due to external time constraints. The limitation of excluding these deaths is addressed in the discussion.

A lifetime QALYs outcome was calculated for each participant by extrapolating the observed within study QALYs variable. Extrapolation was applied by calculating a participant's mean yearly QALYs (total within study QALYs /total years in the study) and applying this mean value to each life year between the study endpoint (age 42) and their final life expectancy. The extrapolation methodology assumed that the distribution of QALYs remains constant with age, as informed through findings from Sullivan et al. (2011).

5.3.3.5 Healthcare Costs

The mean annual per person healthcare costs associated with the five diseases are summarised in Table 5.5. Each mean healthcare cost was derived from published literature: Cancer costs (all cancers across ICD-10 codes C00-D48) were obtained from a study which compared UK research expenditures to total healthcare expenditures (Luengo-Fernandez et al., 2012); costs of depression and obesity were obtained from Rudisill et al. (2016) who used linear regression to identify the direct healthcare costs associated with different BMI categories (18.5-24.9, 25-29.9, 30-34.9, 35-39.9, 40+) and depression status (yes/no); Hex et al. (2012) identified the combined direct healthcare costs for type 1 and type 2 diabetes; and the cost of

hypertension was obtained from a burden of disease model (Lloyd et al., 2003) which identified costs in individual's by blood pressure. In some cases, it was necessary to derive an average annual per person cost by dividing the reported population disease burden by the prevalence of disease. All costs were uprated to 2017/18 prices using the Consumer Price Index published by the ONS (2018b).

Table 5.5: Mean per Person Annual Direct Healthcare Costs by Disease

Disease	Source	Category	Mean Costs
Cancer	Luengo-Fernandez et al. (2012)	Disease (ICD 10: C00-D48)	£2442.49
		No Disease	£0.00
Depression	Rudisill et al. (2016)	Disease	£1095.89
		No Disease	£0.00
Diabetes	Hex et al. (2012)	Disease (Type 1 or Type 2)	£2841.41
		No Disease	£0.00
Hypertension[1]	Lloyd et al. (2003)	Blood Pressure> 140/90 mm Hg	£8.56
		Blood Pressure<= 140/90 mm Hg	£0.00
Obesity[2]	Rudisill et al. (2016)	BMI<=25	£0.00
		25<BMI<=30	£5.24
		30<BMI<=35	£153.26
		35<BMI<=40	£293.92
		BMI>40	£478.66

Notes: Mean costs are per person per year uprated to 2018 prices.
1: Estimates the costs associated with high blood pressure through direct reduction in major cardiovascular events (stroke and ischemic heart failure) for individuals aged 16-44.
2: Estimates costs directly associated with different BMI categories after adjusting the costs of comorbidities including coronary heart disease, stroke, diabetes and cancer.

Costs were assigned to each observation where participants reported a disease between ages of 26-42. Comorbid cases were assigned their largest associated *single* disease cost to avoid double counting. This excluded cases of depression and obesity where comorbid costs could be added as they were obtained from the same regression equation. If no disease was reported it was assumed participants incurred zero healthcare costs. Within study healthcare costs were calculated using the trapezium rule, identifying the area under the curve for all observations between ages 26-42.

Lifetime healthcare costs were estimated by extrapolating within study healthcare costs. It was assumed that a greater proportion of healthcare costs would be accrued by cohort members towards the end of life based on evidence from Asaria (2017) who identifies an increasing distribution of population healthcare costs with respect to age. The population distributions identified by Asaria (2017) were used to estimate the proportion of *lifetime* healthcare costs (p_{hc}) expected to have been accrued by participants during the study period. This proportion was dependent on gender, disease status (at age 42), and life expectancy. Each cohort member's lifetime healthcare costs were estimated by multiplying their aggregated within study healthcare costs by $1/p_{hc}$.

5.3.4 Education Outcome Variables

5.3.4.1 Observed Returns to Education (within study)

Participants' *within study* monetised returns in the education sector were estimated by subtracting their monetary costs incurred from non-compulsory education after the age of sixteen from their monetary returns through economic productivity. This analysis included economic productivity in the construction of educational outcome variable rather than as a separate outcome to avoid double counting. Levin et al. (2017) suggests that many educational polices are justified because of their benefits to future economic productivity, meanwhile, labour market outcomes are often used as a measure of benefit in economic evaluation conducted from the perspective of the education sector (Belfield and Levin, 2007), (Levin and Belfield, 2015).

Costs incurred through additional years of education were assigned by establishing the highest level of education achieved by each participant from self-reported questionnaire items in the BCS 1970 sweep at age 42. Education levels were described as NVQ levels (1-5). No costs were assigned to participants whose highest achievement was NVQ level 1 or 2 as this is equivalent to GCSE/O-levels usually obtained during the period of compulsory education before the age of sixteen. NVQ level 3s, which include A-levels and Scottish Higher Certificates, were assumed to require two years of non-compulsory education and were costed as £4,890 based on standard learner numbers (SLNs). Participants were assumed to be in higher education for three years if they achieved NVQ level 4 (undergraduate degree) or four years if they achieved an NVQ level 5 (post-graduate degree). The average costs of higher education per person per year was estimated as £1500, assuming the public-sector bears 25% of the total costs (total cost per person per year=£6000). All costs were obtained from an econometric study by Hummel et al. (2011) and uprated from 2008/09 prices to 2017/18 prices using the Consumer Price Index (*ONS*, 2018b)

Economic productivity was estimated using the human capital method (Drummond et al., 2015) by accumulating participants' gross earnings over the trial period. All gross earnings were uprated to 2017/18 prices using the Consumer Price Index (*ONS*, 2018b). Accumulated earnings were estimated in three stages. First, the age at which individuals entered the labour market was identified according to the self-reported age at which they left full time education. Participants' earnings upon entering the labour market were estimated as a proportion of their stated earnings at the age 26 sweep. Different proportions were applied for individuals with different NVQ levels. Proportions were calculated based on population data from a study by HM Revenues & Customs (2017) which reports mean earnings for individuals aged 18-25 and 25-30 according to NVQ level.

Second, self-reported gross earnings were established based on self-reports in questionnaires for the five sweeps from ages 26-42. If data was missing due to temporary economic inactivity (e.g. unemployment or temporary sick leave) earnings were imputed as the mean earnings achieved across all BCS 1970 study members by education level (NVQ levels). If data was missing due to permanent economic inactivity (e.g. retirement, looking after family, permanent sick leave) gross earnings were imputed as zero. The trapezium rule was used to estimate accumulated earnings over the study period with observations at labour market entry, and ages 26, 30, 36 and 42.

The third stage accounted for non-accrued earnings due to periods of unemployment. At each sweep participants were asked to report their current and previous economic activities from the date of the last questionnaire – this included all periods of unemployment. Each period of unemployment was multiplied by the mean duration of unemployment for that year as identified by Long (2009). The total reduction to earnings through unemployment was estimated as the total duration of unemployment (years) multiplied by the participant's mean yearly earnings between the ages 16-42. Non-accrued earnings through unemployment were then subtracted from accumulated gross earnings.

5.3.4.2 Extrapolated Lifetime Returns to Education

Lifetime returns to education were estimated by applying distributional assumptions to extrapolate within study gross earnings. As with healthcare costs, the proportion of within study to lifetime earnings (p_e) was not assumed to be equal for all participants. For example, well-educated individuals were in higher education during part of the study period and were therefore expected to accrue a greater proportion of their total lifetime income after the study when compared with participants who had lower levels of education and entered the labour market earlier. Meanwhile, on average, females were likely to have experienced more periods of economic inactivity than males during the study period due to child rearing. Therefore, the value of p_e was dependent on participants' gender and highest achieved NVQ level.

Specific values for p_e were calculated by approximating within study and lifetime earnings in population data published by HM Revenue & Customs (2017). Mean population earnings were summed for each NVQ group across age ranges of 16-42 (to correspond with within study outcomes). This was divided by summed earnings for each NVQ group across age ranges of 16 to retirement age[8] (to correspond with lifetime outcomes), deriving p_e. Lifetime earnings were calculated by multiplying participants' within study earnings by $1/p_e$. The final lifetime returns

[8] Retirement age was assumed to be equal to the age at which individuals could draw a state pension if born in the UK in 1970 (age 65 for men and age 63 for women).

to education outcome was identified by subtracting the previously identified costs of education from participants' lifetime earnings.

5.3.5 Crime Outcome Variables

5.3.5.1 Observed Criminal Justice Costs (within study)

This research identified crime outcomes as public sector costs incurred by the criminal justice sector. Within study crime costs were estimated based on questionnaire responses at ages 30 and 34. Participants were asked to state the total number of times that they had been found guilty of a criminal offence by a court. A public sector cost per criminal conviction was identified as £513.86, which was obtained by uprating the public sector cost per reported crime in studies by Hummel et al. (2011) and Brand and Price (2000) to 2017/18 prices using the CPI (*ONS*, 2018b). The cost per crime did not include any additional personal costs to the victims (e.g. loss of possessions, damages, forgone earnings etc.). Crime costs were identified post analysis by multiplying the public sector cost per crime by the expected number of crimes predicted in the mathematical model. This was necessary as the best fitting regression model (negative binomial) requires outcome variables to be specified as counts (integers).

5.3.5.2 Extrapolated Lifetime Criminal Justice Costs

Lifetime criminal justice costs were estimated by extrapolating participants' within study court convictions to expected lifetime court convictions. A publication by the Home Office (1997) identifies distributions of crime by age and gender and finds most crimes to be committed before individuals are thirty. Extrapolation followed similar methods to those used when estimating lifetime healthcare costs and lifetime earnings. The Home Office (1997) study was used to identify the expected proportion of participants' lifetime crimes occurring during the study period – this proportion was obtained by aggregating population crimes across age groups ranging from 16-34 (the age for which crime data is available in the BCS 1970) and dividing this by the total number of population crimes for all age ranges. The total number of expected lifetime crimes was estimated by multiplying the number of crimes committed by participants at age 34 by the reciprocal of the derived proportion. Different distributions and proportions were identified by gender. Again, costs were applied to lifetime criminal convictions post analysis.

5.3.6 Discounting

Primary (base case) results are reported which discount outcomes (costs and QALYs) at 3.5% as this is the recommended rate in the NICE (2013) reference case. Additional results are presented which discounted outcomes at rates of 1.5% and 0%. Discounting was applied by multiplying the appropriate outcome by a factor of $(1/1 + D)^{n})$ where D is the discount and n is the number of years into the future that the outcomes occur (Severens and Milne, 2004). As the subject of this thesis is concerned with early childhood interventions, it was assumed that discounting should begin at birth – this was appropriate as it corresponds with the time at which a postnatal depression screening intervention would be delivered.

Discounting was applied to the relevant outcome variables within each sweep prior to the implementation of the trapezium rule (which was used to identify aggregated study outcomes). As lifetime outcomes were approximated by multiplying within study outcomes by a scalar, there were no observation points for n after the study period. Therefore, lifetime outcomes occurring after the study period were split into different yearly blocks which ended at final life expectancy for health and crime outcomes or retirement for employment outcomes. Different discount rates were then applied to each yearly block by applying different n values (i.e. n=age 43 for the first yearly block after the study, 44 for the second yearly block etc.).

5.3.7 Empirical Models

5.3.7.1 Specification and Selection of Generalised Linear Models

The results from chapter four suggested that generalised linear models (GLMs) would be an appropriate method to forecast the health and economic lifetime effects from child development measures. GLMs are a class of regression models whose outcome variable (y_i) is assumed to follow an exponential family distribution, and whose mean (μ_i) is assumed to be a function of the independent variable vector (or model input parameters) ($x_i\beta$) (McCullagh, 1984). Each GLM is specified with a specific distribution and a link function which establishes the relationship between μ_i and $x_i\beta$.

The research in this chapter assessed a variety of GLM distribution and link function specifications for each of the outcomes given their differently skewed distributions. These included Ordinary Least Squares (OLS), negative binomial, Poisson, and Gamma regression models. In addition, transformed OLS, two-part, and zero inflated models were assessed to ensure that they did not provide a better prediction of the outcome (Jones and O'Donnell, 2002). All models were estimated using the STATA 12 software.

Several statistical tests/techniques were used to select the most appropriate GLM specification. A set of candidate GLMs was established for each outcome using the modified Park test to identify suitable distributions (Jones, 2010) and the Pregibon test to identify appropriate link functions (Pregibon, 1980). The most appropriate model from all candidate GLMs was then identified as that with the lowest associated Akaike Information Criterion (AIC) (Jones and O'Donnell, 2002). The AIC criterion was also used to compare the fit of GLM models to transformed OLS and two-part regression models. In addition, following recommendation by Desmarais and Harden (2013), a bias corrected Vuong test was used to determine whether zero-inflated Poisson/negative binomial models provided a statistically significant better fit than their GLM counterpart. A more detailed explanation of each statistical test is provided in Appendix 5.2.

Model selection was also influenced by diagnostic plots which determined how closely the empirical data matched the assumptions required for each regression model i.e. deviance residuals are normal and homoscedastic (Barber and Thompson, 2004). Models were rejected if they severely violated these assumptions.

The statistical significance of individual explanatory variables e.g. GCA, RBS, gender, ethnicity etc. was identified using t-tests. The covariates were also grouped into larger categories as either child health, family economic, family psychosocial, neighbourhood or peer influences. Likelihood ratio tests were used to determine the impact of the grouped categories by estimating nested models with/without the grouped covariates (Casella and Berger, 2002). To investigate how outcomes differed across subpopulations, interaction terms were tested between all statistically significant covariates. Interactions were retained if they were statistically significant and they improved the fit of the models as determined through reductions in the AIC statistic (Jones and O'Donnell, 2002).

5.3.7.2 Model Predictions

Model predictions identified the impact per unit increase in GCA/RBS score on all outcome variables. Predictions were obtained on the raw scale of the outcomes for GLM models using the "margins" command in STATA (Williams, 2012). Log transformed OLS models were back transformed manually as exponential retransformation of log models can sometimes result in biased predictions. Bias was accounted for by adding a manually calculated constant, known as Duan's smearing estimator to the regression equation prior to retransformation (Jones, 2010).

To inform the decision model in chapter six, forecasts were obtained estimating the incremental lifetime effects for children exposed to postnatal depression vs. children not exposed.

Incremental lifetime effects were calculated using the results from Growth Curve Model 1 in chapter three, which identified incremental effects in children exposed to postnatal depression symptoms on GCA and SDQ outcomes at age 11.

Lifetime predictions were obtained by multiplying the incremental effect sizes for GCA and SDQ outcomes in chapter three by each models' marginal effect per unit increase in GCA and RBS scores (identified in this chapter). For an illustrative example, assuming the hypothetical incremental effects of postnatal depression from chapter three were GCA=-2 and SDQ=3 and the magnitude per unit increase on lifetime healthcare costs from this chapter equalled -£100 for GCA and £300 for SDQ. Then the incremental effects of postnatal depression exposure on lifetime healthcare costs would be (-2 ×-£100) + (3×£300) = £1,100.

5.4 Results

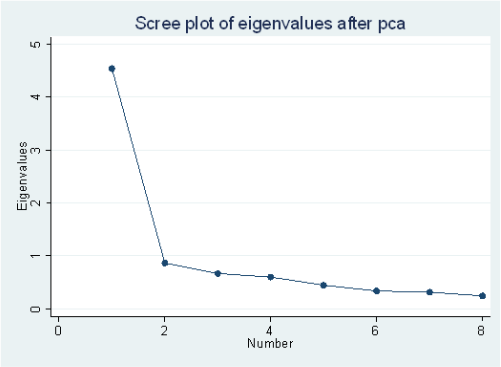
5.4.1 Identifying General Cognitive Ability

A high level of correlation is required between the variables informing Principal Component Analysis (PCA). Correlations were observed between all eight cognitive variables obtained in the BCS 1970, these ranging from 0.29 to 0.73 (Table 5.6).

Principal component analysis was conducted to identify the factors (principal components) which explained the correlation between the cognitive variables in the BCS 1970. The PCA identified a single component with an eigenvalue above one, as is indicated in the scree plot in Figure 5.3. This single factor explained 57% of the total correlation between the eight BCS 1970 cognitive variables. The significant principal component (Component 1 in Figure 5.3) was assumed to represent the underlying cognitive ability of children in the BCS 1970 at age 10 and was suitably renamed General Cognitive Ability (GCA).

Table 5.6: Correlation Matrix between Cognitive Variables

		(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
BAS Word Definitions	(A)	1.00							
BAS Recall of Digits	(B)	0.31	1.00						
BAS Matrices	(C)	0.46	0.29	1.00					
BAS Word Similarities	(D)	0.62	0.30	0.46	1.00				
Pictorial Language Comprehension	(E)	0.64	0.30	0.50	0.56	1.00			
Spelling Dictation Task	(F)	0.43	0.38	0.38	0.38	0.38	1.00		
Edinburgh Reading Test	(G)	0.63	0.38	0.58	0.56	0.62	0.64	1.00	
Friendly Maths Test	(H)	0.60	0.39	0.63	0.56	0.59	0.55	0.73	1.00



Component	Eigenvalue	Total Variance Explained
1	4.53	57%
2	0.86	11%
3	0.67	8%
4	0.60	8%
5	0.44	6%
6	0.34	4%
7	0.31	4%
8	0.24	3%

Figure 5.3: A Scree Plot (left) and Table (right) illustrating the outputs of PCA. Note that Eigenvalues>1 are indicative of important components.

The validity of the PCA assumption (that component 1 represented GCA) was investigated by comparing mean scores for the BCS 1970 GCA variable. The assumption appeared to have face validity as the derived GCA variable followed similar patterns to mean IQ scores for different sub-groups of population i.e. largely equivalent scores for males and female, higher scores for white British children compared with ethnic minorities, and lower scores for children with extremely low birth weights.

5.4.2 Descriptive Analysis

5.4.2.1 Response Rates

The BCS 1970 obtained responses from 16,569 cohort members during the birth sweep. In general, there was a high level of attrition with the overall response number tending to decrease over time (Table 5.7). Only 3,423 (20.66%) of the BCS 1970 cohort provided responses to at least one questionnaire item across *all* nine sweeps (Mostafa and Wiggins, 2014).

Table 5.7: Total Number of Questionnaire Responses in BCS 1970

Sweep Number	Age	Response (n)[1]	Response (%)[2]
1	Birth	16,569	100.00
2	5	12,939	78.09
3	10	14,349	86.60
4	16	11,206	67.63
5	26	8,654	52.23
6	30	10,833	65.38
7	34	9,316	56.23
8	38	8,545	51.57
9	42	9,354	56.45
All Sweeps	N/A	3,423	20.66

Notes: Table adapted from Mostafa and Wiggins (2014) pp.7-8
1: Responses counted if cohort members answer one or more questionnaire item(s) within a sweep
2: Response % is taken from number responding to the initial birth survey

There were relatively low response rates for QALYs (n=5,111), healthcare costs (n=4,478), and returns to education (n=3,772) as these outcome variables were both derived using multiple questionnaire items and accumulated using data from multiple sweeps. A larger number of cohort members (n=8,864) provided a response to the question in the age 34 sweep relating to their total number of criminal convictions.

Both General Cognitive Ability and the Rutter Behaviour Scale variables were associated with relatively high levels of missing data which is likely due to being multi-attribute questionnaires/ being derived from multiple psychometric tests. Generally, less missing data occurred in the other explanatory variables as these were obtained during childhood sweeps where response rates were highest. Full response rates across all variables are reported in Table 5.8. The limitations associated with missing data are discussed in section 5.5.4.

Table 5.8: Response Numbers across Outcome & Explanatory Variables

	Age	Response (n)[1]	Response (%)[2]
Outcome Variables			
QALYS	16-42	5,111	30.85%
Healthcare Costs	16-42	4,478	27.03%
Returns to Education	16-42	3,772	22.77%
Criminal Convictions	34	8,864	53.50%
Explanatory Variables			
General Cognitive Ability	10	8,102	48.90%
Rutter Behaviour Scale	10	11,236	67.81%
Ethnicity	Birth	13,712	82.76%
Sex	Birth	14,984	90.43%
Hypertension	10	12,520	75.56%
BMI	10	12,159	73.38%
Cumulative Health Problems	10	11,907	71.86%
Family Income	10	12,508	75.49%
Age Father Left Education	Birth	16,275	98.23%
Stable Family	10	12,791	77.20%
Large Family	10	13,636	82.30%
Crowding in Home	10	12,943	78.12%
Freq. Family Activities	10	13,329	80.45%
Neighbourhood Quality	5	12,711	76.72%
Peers Smoke	10	12,564	75.83%

Notes: 1: Responses counted if cohort members answer all questionnaire item used to derive the variable
2: Response % is taken from number responding to the initial birth survey

5.4.2.2 Distribution of Outcomes

All outcome variables appeared to follow non-normal distributions which is often the case when dealing with variables that relate to count data and monetary costs. Healthcare costs and total number of crimes had positively skewed distributions while a negative skew was observed for QALYs as HRQoL *decrements* were assigned to diseases participants, Figure 5.4.

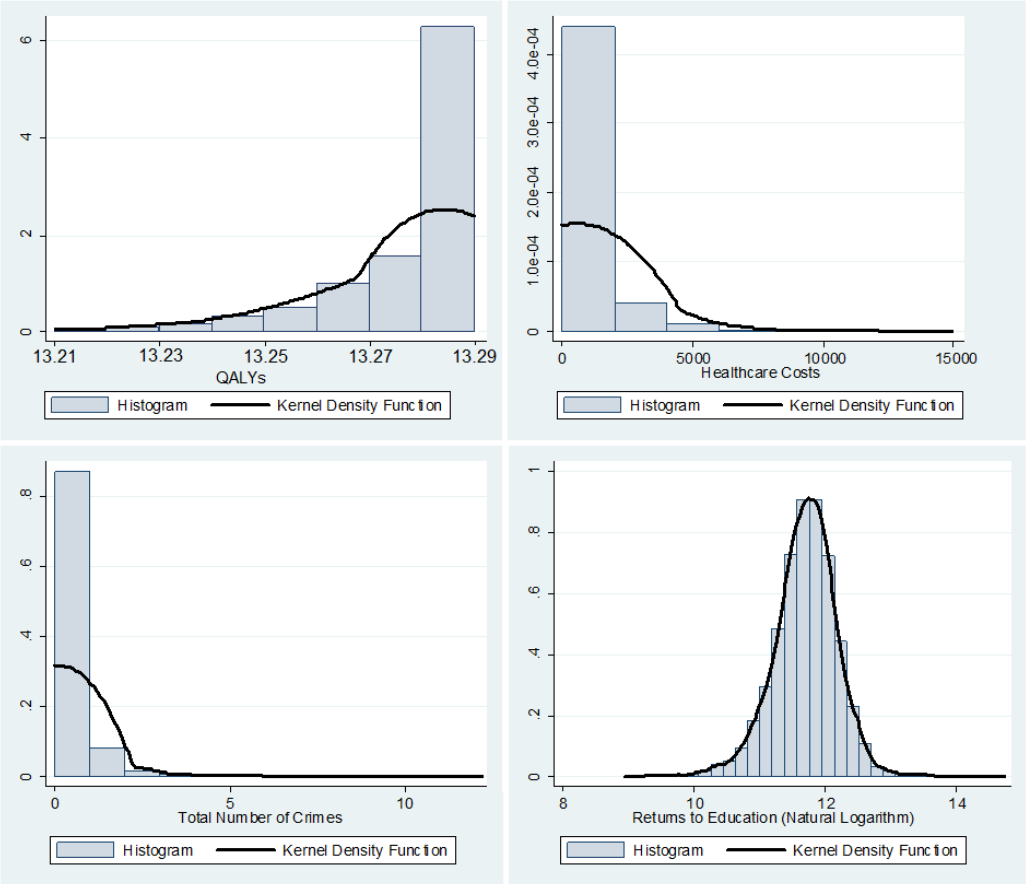


Figure 5.4: Illustrates the distribution of outcome variables using histograms overlaid with kernel density plots. Outcomes are within study QALYs (top left), healthcare costs (top right), total number of court convictions (bottom left) and the natural logarithm of returns to education (bottom right).

Both health outcomes had a large proportion of null observations as many individuals did not suffer from any of the five diseases. This resulted in most participants being assigned zero values for healthcare costs and mean QALYs for non-diseased individuals (equal to 13.28). All participants had positive returns to education which closely resembled a normal distribution after a log-transformation was applied to the outcome variable (Figure 5.4).

5.4.2.3 Associations between Child Development Measures and Outcomes

Associations were observed between participants GCA at age 10 and mean/median scores on each adulthood outcome. This association followed a gradient where the highest (lowest) scoring GCA quartiles were associated with the most (least) desirable outcomes, Table 5.9. Similarly, participants who scored in the best (worst) percentile on the Rutter Behaviour Scale (RBS) were associated with the most (least) desirable outcomes for QALYs, healthcare costs and crime costs. There did not appear to be an associated between RBS scores and median returns to education, Table 5.9.

Table 5.9: **Median/Mean Outcomes by Child Development Percentile**

	QALY[1]	Healthcare Costs[1]	Returns to Education[1]	Crime Costs[2]
General Cognitive Ability				
Lowest 25% (worst performers)	12.997	£348.44	£306,800	£161.17
25th- 50th Percentile	13.026	£83.84	£346,400	£149.23
50th – 75th Percentile	13.138	£73.36	£360,000	£114.24
75th – 100th Percentile	13.138	£52.40	£422,900	£62.79
Rutter Behaviour Scale				
Lowest 25% (best performers)	13.138	£62.88	£369,600	£85.16
25th- 50th Percentile	13.138	£73.36	£363,300	£107.82
50th – 75th Percentile	13.099	£73.36	£366,100	£117.51
75th – 100th Percentile	13.026	£337.96	£363,400	£170.24

Notes: All are undiscounted within study outcomes. 1=median and 2=mean

5.4.2.4 Associations between Covariates and Outcomes

The association between model covariates and outcomes are reported in Table 5.10. Adulthood health was strongly associated with childhood health, where median (undiscounted within study) QALYs were lower and healthcare costs were larger for: the highest childhood BMI quartiles (12.86, £971.96) compared with the lowest (13.19, £31.44); for those with hypertension during childhood (12.91, £112.84) compared with normal blood pressure (13.10, £73.36) and for participants who suffered from a high number of diseases (12.68, £379.88) versus participants with a history of few disease (13.10, £73.36). Health outcomes also appeared to be worse for participants who were not ethnically white British, and who came from a large family with low income.

Gender was strongly associated with mean (undiscounted within study) crime costs during adulthood (male =£232.50/ female=£33.11), as were family income (highest quartile =£97.06/lowest quartile= £237.27), neighbourhood quality (good quality= £87.63/ poor quality =£210.60), being from a stable family (yes= £108.95/ no= £202.26) and the smoking status of participants' peer groups (smoke= £206.34/don't smoke= £112.30).

Median (undiscounted within study) returns to education were much larger for males (£446,700) than for females (£287,900); for participants who came from high income families (highest quartile=£400,600/lowest quartile=£350,100) whose fathers were highly educated (highest quartile= £398,700/lowest quartile= £338,900), and for participants from good neighbourhoods (good neighbourhood=£396,500/ poor neighbourhood=£358,900) whose peers did not smoke (peers smoke=£399,00/ peers do not smoke=£360,300). There were no obvious associations between returns to education and childhood ethnicity, child health and family psychosocial variables, Table 5.10.

Table 5.10: Median/Mean Outcomes by Covariate Category

	QALYs[1]	Healthcare Costs[1]	Returns to Education[1]	Crime Costs[2]
Child Demographic Covariates				
Gender				
Male	13.138	£73.36	£446,700	£232.50
Female	13.026	£65.68	£287,900	£33.11
Ethnicity				
White British	13.099	£73.36	£365,900	£125.98
Not White British	12.997	£83.84	£374,800	£174.54
Child Health Covariates				
BMI				
Lowest (best) 25%	13.188	£31.44	£368,400	£125.89
25-50th Percentile	13.138	£62.88	£391,400	£134.93
50th-75th Percentile	13.138	£83.84	£378,600	£137.01
Highest 25%	12.855	£971.96	£345,200	£105.13
Hypertension				
No	13.099	£73.36	£366,500	£128.88
Yes	12.906	£112.84	£365,400	£61.27
>=5 Childhood Diseases				
No	13.099	£73.36	£368,500	£121.77
Yes	12.680	£379.88	£325,700	£114.47
Family Economic Covariates				
Family Income				
Lowest (worst) 25%	13.048	£83.84	£350,100	£237.27
25-50th Percentile	13.026	£90.48	£344,600	£142.53
50th-75th Percentile	13.099	£73.36	£356,800	£118.01
Highest 25%	13.138	£62.88	£400,600	£97.06
Age Father Left Education				
Lowest (worst) 25%	13.026	£81.40	£338,900	£125.35
25-50th Percentile	13.032	£83.84	£349,900	£149.08
50th-75th Percentile	13.138	£62.88	£382,900	£90.63
Highest 25%	13.138	£52.40	£398,700	£73.74
Family Psychosocial Covariates				
Stable Family				
No	13.026	£90.48	£372,200	£202.26
Yes	13.138	£73.36	£364,600	£108.95
Large Family				
No	13.099	£73.36	£368,000	£123.53
Yes	13.008	£414.12	£320,000	£181.36
Ratio of people to rooms				
Lowest (best) 25%	13.138	£73.36	£382,400	£75.10
25-50th Percentile	13.099	£73.36	£363,500	£111.43
50th-75th Percentile	13.138	£62.88	£374,500	£150.56
Highest 25%	13.048	£83.84	£349,000	£165.61
Number of Family Activities				
Lowest (worst) 25%	12.855	£83.84	£323,400	£307.44
25-50th Percentile	13.096	£112.84	£317,500	£224.95
50th-75th Percentile	13.026	£83.84	£367,400	£164.64
Highest 25%	13.046	£73.36	£364,400	£132.81
Neighbourhood and Peer Covariates				
Neighbourhood Quality				
Poor	13.006	£102.36	£358,900	£210.60
Average	13.099	£73.36	£356,700	£130.67
Good	13.138	£62.88	£396,500	£87.63
Peer Smoking				
No	13.099	£73.38	£360,300	£112.30
Yes	13.096	£83.84	£399,000	£206.34

Notes: All are undiscounted within study outcomes. 1=median and 2=mean

5.4.2.5 Extrapolation and Discounting

The effect of extrapolation was investigated by comparing (mean and median) within study outcomes with lifetime counterparts (all discounted at 3.5%). The extrapolated lifetime outcomes were considerably larger than the observed within study outcomes for variables in the health and education sectors. For example, median within study vs lifetime outcomes equalled 4.13 vs 7.62 QALYs, £22.12 vs £44.07 for healthcare costs and £120,000 vs. £190,000 for returns in the education sector. In contrast mean lifetime crime costs (£75.88) were only slightly larger than the mean crime costs observed within the study (£69.83), Figure 5.5.

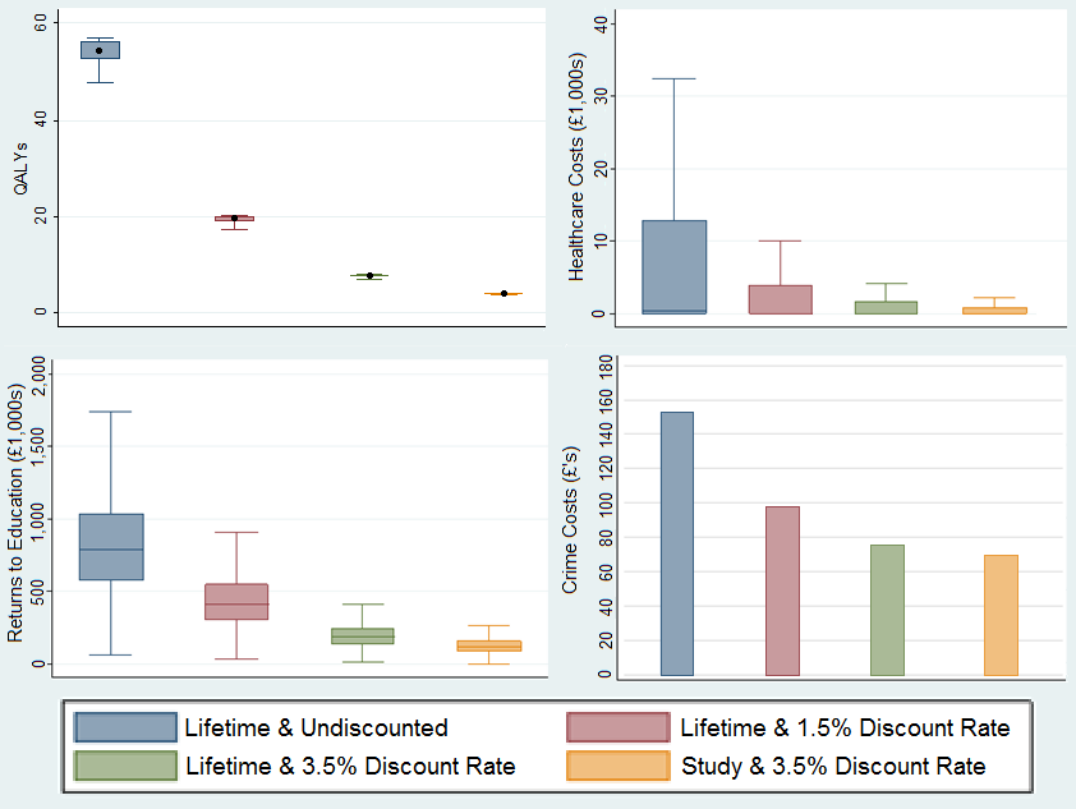


Figure 5.5: Box and whisker plots illustrating median values and inter-quartile ranges for QALYs (top left), healthcare costs (top right), and returns to education (bottom left). A bar graph reports mean crime costs (bottom right). Plots are obtained according to discount rate and for within study and lifetime outcomes.

Similarly, mean/medians were identified to establish the impact of discounting. The application of different discount rates had a substantial effect on the size and the variability of all outcomes. Again, these effects were larger for health and education, for example median lifetime undiscounted outcomes (QALYs=54.20; healthcare costs=£376.16; returns to education=£790,000) were over four times the size of median lifetime outcomes discounted at 3.5% (QALYs=7.62; healthcare costs=£44.07; returns to education=£190,000). As most crimes occurred earlier in the lifespan, the impact of discounting was less severe with mean lifetime undiscounted costs equal to £152.73 compared with mean lifetime costs discounted at 3.5% equal to £75.88, Figure 5.5.

5.4.3 Regression Models

5.4.3.1 Model Specifications

A generalised linear model specified with a log link function and a gamma distribution was used when modelling healthcare costs (n=1,305) – this GLM specification was identified as most appropriate by the modified Park test and Pregibon's link test and outperformed an OLS model which violated assumptions that residuals were normal and homoscedastic. While healthcare costs had a large proportion of zero observations, there was no evidence to suggest that a two-part model (logit and GLM) was more appropriate than the single GLM given the higher associated AIC criterion. In contrast, the QALYs (n=1,488) outcome was modelled using OLS regression as the modified Park and Pregibon's link tests suggested a Gaussian distribution with an identity link function as most appropriate. These results were supported by the lower associated AIC criterion for the OLS model when compared with all other GLM specifications.

Returns to education (n=1,317) were estimated using OLS regression whilst applying a log transformation to the dependent variable. The distribution of the log transformed variable (Figure 5.3, section 5.4.2.1) appeared to follow a normal distribution, and produced a much better fit than OLS regression on the non-transformed variable. Despite the benefit of providing easy to interpret coefficients, GLMs with non-transformed outcomes were rejected as the modified Park test suggested no distribution would provide a suitable fit. This choice was supported by lower AIC associated with the log transformed model when compared to the best fitting GLM.

A GLM with a log link function and a negative binomial distribution was used to model the total number of criminal convictions in 2,869 participants. The negative binomial distribution was selected in favour of the Poisson distribution given the lower associated AIC. This decision was further justified as the data appeared to be over dispersed, meaning the expected variance was greater than the expected mean, thus violating a required assumption in Poission regression of equivalence between expected means and variance. Whilst most participants reported zero values (no criminal convictions) the zero-inflated negative binomial model was not a significant improvement on the selected model according to the bias corrected Vuong test.

5.4.3.2 Model Fit

Figure 5.6 depicts the residual plots for each of the models. The returns to education model appeared best fitting with heteroscedastic residuals that followed a (symmetrically) heavy

tailed distribution mimicking the normal distribution reasonably closely. The QALYs model had heteroscedastic residuals which diverged from normality, displaying a slight positive skew.

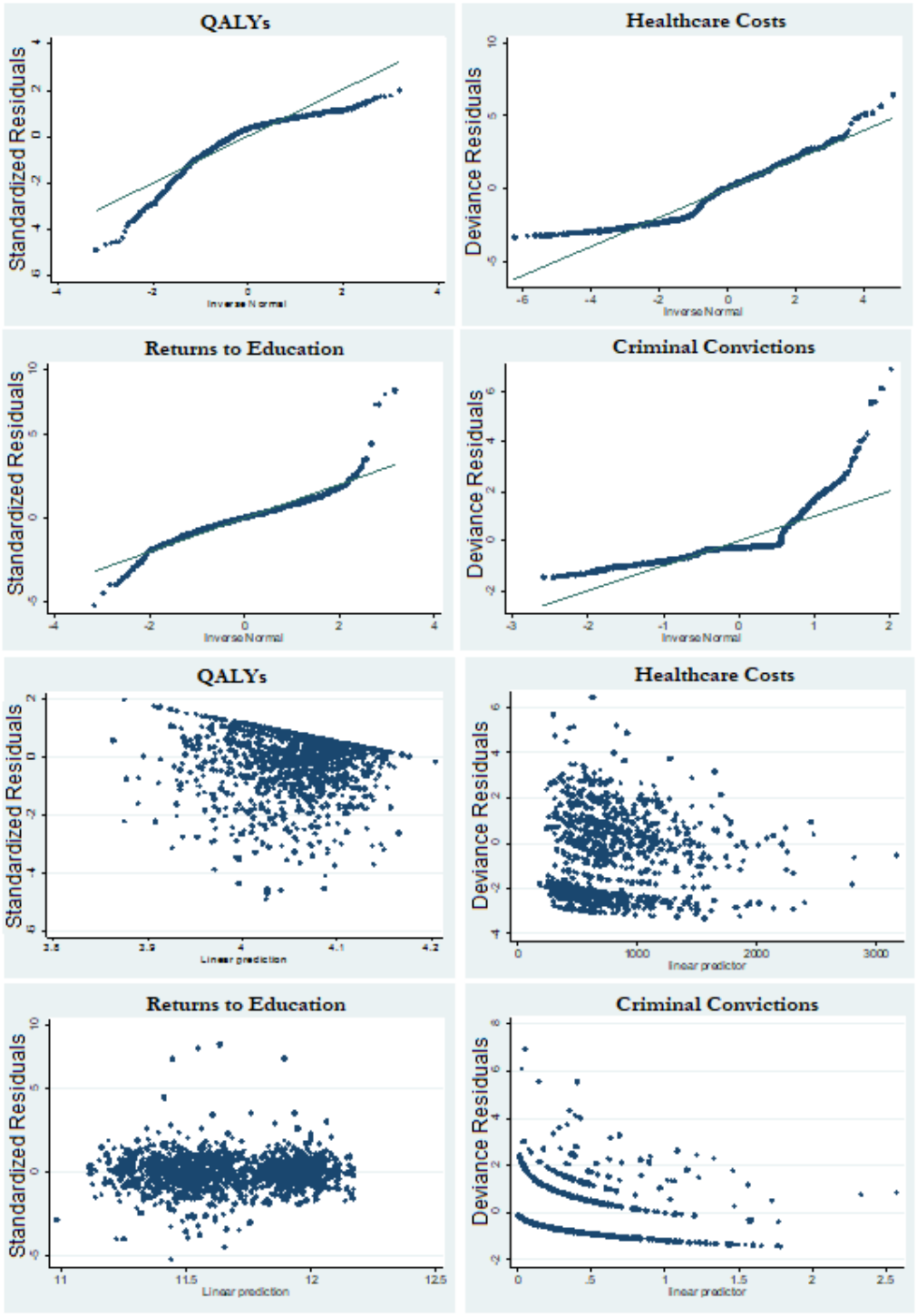


Figure 5.6: Above are Q-norm plots of deviance/standardised residuals, below displays scatter plots of deviance/standardised residuals vs. model predictions for all outcome variables.

The deviance residuals for the healthcare costs model diverged from normal at the tail of the distribution and were largely heteroskedastic but may have demonstrated a slight tendency to

decrease for larger healthcare costs. The worst fitting model appeared to be the negative binomial regression model predicting participants' total number of crimes, where the deviance residuals appeared to be neither normal nor heteroscedastic, Figure 5.6. The importance of violations to model assumptions is discussed in detail later in the chapter.

5.4.3.3 Statistical Inference

All models revealed strong and consistent associations between measures of child development and adulthood outcomes. Increased GCA (better cognitive outcomes) at age 10 was significantly associated with increased QALYs ($p<0.0001$), reduced healthcare costs ($p=0.015$), positive returns to education ($p=0.011$), and fewer criminal convictions ($p=0.002$). Increased RBS (worse socioemotional outcomes) at age 10 was significantly associated with reduced QALYs ($p=0.017$), increased healthcare costs ($p=0.008$), and more criminal convictions ($p=0.002$), and was non-significantly associated with reduced returns to education ($p=0.472$), Table 5.11.

There was also an association between all outcomes and gender, where females had significantly worse health outcomes ($p=0.001$ QALYs, $p=0.001$ healthcare costs), lower returns to education ($p<0.0001$) and fewer court convictions ($p<0.0001$) compared with males. Other covariates relating to the participant included: child health which was significantly associated with adulthood health (QALYs & healthcare costs $p<0.0001$) particularly through the effects of childhood BMI; meanwhile a participant's ethnicity was only associated with criminal convictions ($p=0.004$), Table 5.11.

The impact of early environmental covariates differed across outcomes. The most consistent associations were identified for children's neighbourhoods where the most desirable outcomes were more likely to be observed for children growing up in good quality neighbourhoods; these associations were statistically significant for both healthcare costs ($p<0.0001$), and crime costs ($p<0.0001$). The family economic environment was not associated with health outcomes but was associated with returns to education ($p=0.0073$) and the number of court convictions ($p<0.0001$). Meanwhile the family psychosocial environment was significantly associated with healthcare costs ($p<0.0001$) and criminal convictions ($p<0.0001$) but not QALYs or returns to education.

All hypothesis tests are reported for within study outcomes discounted at 3.5%. These interpretations remained constant irrespective of whether outcomes were discounted at different rates or whether the time horizon was for the study period or extrapolated over the lifetime. This excluded the effects of gender on lifetime QALYs which were influenced by the additional life expectancy experienced by females.

Table 5.11: Untransformed Model Coefficients and Associated P-values

	QALYs			Healthcare Costs			Returns to Education			Crime Costs		
	Coeff.	P-Value (t-test)	P-Value (lr-test)	Coeff.	P-Value (t-test)	P-Value (lr-test)	Coeff.	P-Value (t-test)	P-Value (lr-test)	Coeff.	P-Value (t-test)	P-Value (lr-test)
Child Development Variables			<0.0001			<0.0001			0.0092			<0.0001
General Cognitive Ability	0.0015	<0.0001		-0.0110	0.0150		0.0037	0.0110		-0.0158	0.0020	
Rutter Behaviour Scale	-0.0115	0.0170		0.1948	0.0080		-0.0113	0.4720		0.2540	0.0020	
Child Demographic Variables						0.0010			<0.0001			<0.0001
Female	-0.0281	0.0010		0.3903	0.0010		-0.4932	<0.0001		-2.2135	<0.0001	
White British	0.0479	0.0680		0.0362	0.8950		0.1731	0.0530		-0.8666	0.0040	
Child Health Variables			<0.0001			<0.0001			N/A			N/A
Hypertension	-0.0128	0.5510		-0.3092	0.2970		N/A	N/A		N/A	N/A	
BMI	-0.0166	<0.0001		0.1449	<0.0001		N/A	N/A		N/A	N/A	
Cumulative Health Problems	-0.0927	0.001		0.3996	0.1870		N/A	N/A		N/A	N/A	
Family Economic Variables			0.7338			0.4009			0.0073			<0.0001
Log Family Income	-0.0039	0.6930		0.0589	0.6420		0.0799	0.0100		0.1838	0.2410	
Age Father Left Education	0.0012	0.4580		-0.0156	0.5620		0.0068	0.1960		-0.0823	0.0240	
Family Psychosocial Variables			0.5165			<0.0001			0.0626			<0.0001
Stable Family	0.0054	0.7010		-0.1093	0.5280		-0.0925	0.0430		-0.1792	0.2950	
Large Family (>=4 Siblings)	-0.0348	0.1790		0.3931	0.2510		-0.1395	0.1130		-0.8059	0.0220	
Crowding (people/rooms)	0.0223	0.1940		-0.007	0.9980		0.0687	0.2580		N/A	N/A	
Freq. Family Activities	-0.0017	0.7400		0.1568	0.0150		0.0220	0.1870		0.0454	0.5100	
Neighbourhood & Peer Variables			0.4366			<0.0001			0.6028			<0.0001
Poor Neighbourhood	0.0042	0.8550		0.2449	0.4400		0.0444	0.5700		0.7069	0.0050	
Good Neighbourhood	0.0120	0.2020		-0.0955	0.5080		0.0363	0.2320		0.7069	0.0460	
Peers Smoke	N/A	N/A		N/A	N/A		0.0200	0.6420		-0.0825	0.6530	
Interaction Terms												
GCA#Female	N/A	N/A		N/A	N/A		0.0060	0.0030		N/A	N/A	

Notes: Reports untransformed β coefficients and the associated p-values for all explanatory variables. P-values obtained from t distribution for individual covariates, and from likelihood ratio tests for groups of covariates. Results are for within study outcomes discounted at 3.5%.

5.4.3.4 Model Predictions

Model predictions on the raw scale of the outcome variables are reported in Table 5.12. For within study outcomes discounted at 3.5%, a unit increase in a participant's GCA score at age 10 (better cognitive development) predicted: an increase in QALYs of 0.001; decreases in healthcare costs of -£8.04; £2,514.17 accrued through additional returns to education and small reductions in costs incurred by the crime sector (-£0.92). Meanwhile, a one unit increase in a participant's standardised Rutter Behavioural Scale (RBS) score (worse socioemotional development) predicted: a 0.012 decrease in QALYs; an increase in healthcare costs of £143.20; negative returns to the education equal to -£3,166.20 and additional crime costs of £14.84.

Table 5.12: Mean Marginal Effects of GCA/RBS by Time Horizon and Discount Rate

	Within Study (Discount =3.5%)	Lifetime (Discount =3.5%)	Lifetime (Discount =1.5%)	Lifetime (Discount = 0 %)
QALYs				
GCA	0.001 (0.001, 0.002)	0.003 (0.001, 0.004)	0.007 (0.004, 0.010)	0.019 (0.010, 0.027)
RBS	-0.012 (-0.021, -0.002)	-0.021 (-0.039, -0.004)	-0.055 (-0.101, -0.010)	-0.155 (-0.282, -0.027)
Healthcare Costs				
GCA	-£8.04 (-£14.50, -£1.63)	-£15.81 (-£28.96, -£2.66)	-£40.62 (-£75.40, -£5.85)	-£133.58 (-£249.68, -£17.49)
RBS	£143.20 (£35.50, £250.92)	£292.42 (£72.71, £512.12)	£769.87 (£190.32, £1349.42)	£2564.39 (£633.47, £4495.30)
Returns to Education[1]				
GCA	£2,514.17 (£1,611, £3,419)	£4,524.81 (£3,153, £5,900)	£10,895.39 (£7,840, £13,958)	£21,781.50 (£15,914, £24,690)
RBS	-£3,166.20 (-£16,064, £10,181)	-£3,104.10 (-£22,741, £17,206)	-£4,342.77 (-£48,130, £40,965)	-£5,312.39 (-£89,453, £81,779)
Crime Costs				
GCA	-£0.92 (-£1.55, -£0.30)	-£1.00 (-£1.68, -£0.32)	-£1.29 (-£2.16, -£0.42)	-£2.02 (-£3.38, -£0.65)
RBS	£14.84 (£4.91, £24.78)	£16.13 (£5.33, £26.92)	£20.72 (£6.85, £34.58)	£32.46 (£10.73, £54.19)

Notes: Mean (and 95% confidence intervals) reported for all outcomes. Each marginal effect is per unit increase in GCA and standardized RBS.
1: Results were back transformed to the raw scale using Duan's smearing estimator.

There was a substantial difference between within study/lifetime predictions and between outcomes discounted at different rates (Table 5.12). This was most evident when comparing healthcare costs where the predicted size per unit increase in standardised RBS was over seventeen times larger for undiscounted lifetime outcomes (£2564.39) when compared to within study outcomes discounted at 3.5% (£143.20). Similar effects were observed when predicting returns to education where each unit increase in GCA was over eight times larger for

undiscounted lifetime outcomes (£21,781.50) vs. within study outcomes discounted at 3.5% (£2,514.17).

5.4.4 Lifetime Incremental Effects for Children Exposed to Postnatal Depression

Incremental mean *lifetime* outcomes, discounted at 3.5%, were estimated for children exposed to maternal depression symptoms using results from chapter three, from Growth Curve Model 1. As the effect sizes from Growth Curve Model 1 were identified on the SDQ total difficulties scale, results were obtained for each model in this chapter using an equivalent measurement scale where the standardised RBS was transformed to a scale ranging from 0-40. The full results for the RBS transformed scale are described in Appendix 5.3.

Children exposed to symptoms of postnatal distress at nine months were predicted to have worse cognitive and socioemotional outcomes, with incremental effects equal to -0.27 (GCA) and 2.07 (SDQ). The incremental effects on child development measures were predicted to translate into fewer lifetime QALYs (mean=-0.01), increased lifetime healthcare costs (£125.61), fewer lifetime returns to the education sector (-£2,512.52), and increased lifetime costs to the criminal justice sector (£6.96) when compared with children whose mothers were not depressed during the postnatal period, Table 5.13a & 5.13b.

Table 5.13a: Incremental Effects of Maternal Depression on Child Development

	GCA	SDQ/ Rutter
Postnatal depression symptoms[1]	-0.27	2.07
Maternal depression symptoms[2]	-1.83	1.19

Notes: Results obtained from chapter three, Growth Curve Model 1.
Incremental effects vs. children not exposed to maternal depression symptoms.
1: Children exposed to postnatal depression symptoms at 9 months.
2: Per episode of exposure to maternal depression symptoms for children aged 3-11.

Table 5.13b: Incremental Effects of Maternal Depression on Lifetime Outcomes

	Adulthood Effects (Lifetime, Discount=3.5%)			
	QALY s	Healthcare Costs	Returns to Education	Crime Costs
Postnatal depression symptoms[1]	-0.01	£125.61	-£2,512.52	£6.96
Maternal depression symptoms[2]	-0.01	£98.86	-£9,024.31	£5.69

Notes: Incremental effects vs. children not exposed to symptoms of maternal depression.
1: Children exposed to postnatal depression symptoms at 9 months.
2: Per episode of exposure to maternal depression symptoms for children aged 3-11.

The impact per each additional exposure to maternal depression symptoms occurring between ages three and eleven was identified using the cumulative distress variable from chapter three. Incremental effects were like those of postnatal depression and resulted in fewer QALYs (- 0.01), additional healthcare (£98.86) and crime (£5.69) costs, and fewer returns to education (-£9,024.31) across the lifespan.

The results predicted large lifetime effects (discounted at 3.5%) for children repeatedly exposed to maternal depression throughout their childhood. These were obtained from chapter three, Growth Curve Model 2 where the incremental difference between children exposed to postnatal symptoms and persistent symptoms of depression (vs. no symptoms) were substantial (-0.04 QALYs, £473.46 healthcare costs, -£30,740.97 returns to education, £26.86 crime costs). Full results for trajectories from Growth Curve Model 2 are reported in Appendix 5.4.

The lifetime effects estimates can be used to establish the overall population disease burden associated with postnatal depression. Assuming a prevalence rate of depression as 13% per birth (Leahy-Warren and McCarthy, 2007), the burden of postnatal depression (through effects to children) for a one-year cohort of 774,835 live births in the UK for 2016 (*ONS*, 2017) is projected to equal 916 QALYs, £12.6 million through direct healthcare costs, £253.5 million in the education sector, and £702,000 in crime costs. If it is assumed that roughly 30% of postnatally depressed women suffer from at least one further episode of depression symptoms later during childhood (Vliegen et al., 2014) the impact of maternal depression on a one-year cohort of children is estimated to result in *at least* 1,218 fewer QALYs, £15.6 million in additional healthcare costs, £526 million fewer returns in education, and £900,000 in crime costs.

5.5 Discussion

5.5.1. Key Findings

This analysis identified a strong association between cognitive and socioemotional development at age 10 and later life outcomes relevant as decision endpoints for economic evaluation. The empirical results support developmental theory and suggest that improvements to children's cognitive performance and reductions to their behavioural and emotional problems are likely to result in improved health outcomes, reduced healthcare costs, increased returns to education, and reduced costs to the criminal justice sector during adulthood.

Results were wholly consistent, occurring in the expected direction, with both cognitive and socioemotional development measures associated with later life outcomes and seven of eight

outcomes being statistically significant. The magnitude of effects was large for the returns to education outcome, not trivial for later life QALYs and healthcare costs, but insubstantial for crime costs. These key findings are supported by similar results from empirical studies in UK cohorts linking measures of child development at (roughly) age 10 to self-rated health (Layard et al., 2014), (Hertzman et al., 2001), employment/education (Layard et al., 2014), (Healey et al., 2004), and criminal outcomes (Herrnstein and Murray, 2010).

5.5.2 Informing Policy through Model Predictions

The models presented in this analysis provide a mechanism for estimating the lifetime effects of childhood health interventions, if evidence is available regarding the incremental effects associated with the intervention on General Cognitive Ability (β_1) and/or the Rutter Behaviour Scale (β_2) at age 10. Given this availability, mean (per individual) QALYs, healthcare costs, monetised returns to education and costs of crime can be estimated by multiplying the GCA and Rutter Behaviour Scale regression coefficients reported in this analysis by factors of β_1 and/or β_2.

The estimation of lifetime effects could be useful for decision makers when assessing the cost effectiveness of childhood interventions. As was described in Chapter 4, no other models were identified that obtained effect sizes on multiple decision endpoints directly relevant for economic evaluations conducted in the UK (i.e. QALYS and healthcare costs).

Additionally, this study was tailored for UK evaluations, obtaining estimates for the lifetime horizon and with discount rates at 0%, 1.5% and 3.5% in accordance with recommendations in the NICE (2013) reference case. The results for each model and for each discount were reported separately, providing a useful source for analysts who might wish to apply different discount rates to healthcare costs and benefits (QALYs).

5.5.3 Lifetime Effects of Postnatal Depression

The above methodology was applied in this research to estimate the lifetime effects for children exposed to symptoms of postnatal depression using incremental effect sizes obtained from the analysis in chapter three. The estimation of lifetime effects demonstrates the full indirect estimation process and fulfils one of the overarching aims of the research in this thesis. The predicted incremental difference in lifetime effects for children exposed to postnatal depression symptoms vs. not exposed were not trivial and may have implications for health decision making. These estimates are used in chapter six to complete the applied example by assessing

the cost-effectiveness of screening for postnatal depression whilst considering effects to children, updating a NICE (2018) report which did not include child outcomes.

Whilst the predicted magnitude of lifetime effects for each individual child exposed to symptoms of postnatal depression appeared relatively modest, the predicted lifetime effects of postnatal depression on population outcomes were substantial (916 QALYs, £12.6 million through direct healthcare costs, £253.5 million in the education sector, and £702,000 in crime costs). The yearly disease burden of postnatal depression through effects to children was large given the high population prevalence of postnatal depression and the subsequent high risk of childhood exposure to symptoms (McCarthy 2007). These lifetime effects to children are in addition to the detrimental health and social effects incurred by mothers who suffer from depression symptoms.

5.5.4 Limitations

The results from this research should be interpreted with caution given several limitations in this empirical analysis. Firstly, the fit between the models and the empirical data was not optimal. Regression assumptions that (deviance) residuals were homoscedastic and normal were not met for the crime outcome and appeared to be somewhat violated for both QALYs and healthcare costs outcomes. It is possible that these violations may have led to imprecise and potentially biased estimates (Dobson and Barnett, 2008).

The cause of heteroscedastic (deviance) residuals may have resulted from omitted/unobserved covariates. For example, concurrent adulthood variables (e.g. adulthood income, marital status and health behaviours) likely to be associated with the outcomes were not included in the analysis. This was decided as theory implies that child development is likely to predict both adult concurrent and outcome variables (Overton, 2003). Therefore, the inclusion of concurrent adulthood variables in a GLM may have resulted in multicollinearity and reduced the interpretability of regression coefficients attached to the child development measures (Farrar and Glauber, 1967). Evidence from a similar empirical study suggests the relative impact of excluding adulthood variables may have been insubstantial: Hertzman et al. (2001) predict adulthood self-rated health using GLMs and find consistent associations for cognitive and socioemotional development in the presence and absence of concurrent adulthood variables.

Nonetheless, results from this analysis might be improved using a more sophisticated model which could be estimated using structural equation regression methods. This type of model could account for the effects of child development on lifetime outcomes which occur through their effects on concurrent adulthood outcomes. While this was beyond the scope of this chapter it should be an objective for future research.

For health outcomes, model violation may have occurred due to the distribution of the outcome variables. No direct questionnaire items were available to identify healthcare costs or HRQoL, so participants were assumed to incur the mean (per person) HRQoL decrement and healthcare costs associated with five diseases (cancer, depression, diabetes, hypertension or obesity). As some values occurred at a relatively high frequency (e.g. obesity), the distribution of the outcome variable may have been multimodal despite there being a possible 30^4 unique values. One of the limiting factors of working with large cohort data is the reliance on measurements not pre-specified by the investigators. It would be very useful for health economics research if future cohort studies prioritised the collection of healthcare cost and HRQoL data.

The measures available in the BCS 1970 might have led to further research limitation, as the model outcome may not have identified all the important decision endpoints. For instance, it was not possible to include within study deaths in the calculation of within study QALYs as this information required a special user's licence. According to a report by Heywood et al. (2015) roughly 1% of participants died during the period for which QALYs were obtained in this study.

Additional health effects would also have occurred in BCS 1970 study participants through diseases beyond the five for which data was available – these effects are no less relevant to decision makers. Furthermore, all costs of crime were unlikely to be identified from information on court convictions as costs occur in other areas for example, extending to victims through the crimes themselves e.g. theft, vandalism etc. (Brand and Price, 2000). Although it is impossible to know the exact impact of the unobserved outcomes on model interpretations, it might be sensible to assume that effect sizes occur in the same direction as the observed outcomes. For example, the impact of child development on, say, coronary heart disease may well be similar to the association between child development and cancer, depression, diabetes, obesity and hypertension. If this were the case, then predictions from this analysis would conservatively estimate the lifetime effects of child development on lifetime outcomes.

Effect sizes on health outcomes may have been doubly underestimated given the assumptions used to calculate life expectancies from cohort members' disease status at age 42. Firstly, most of the cancer, diabetes and hypertension cases may be expected later in life but the extrapolation did not predict additional cases of disease with onset after the study endpoint. Secondly, the % life years lost per disease were obtained from an NHS Executive (1996) report, but these calculations were not stratified by age of disease onset. On average, individuals who experience diseases earlier in life would have a higher % life years lost than individuals with later onset because they have a larger proportion of life expectancy to lose. Therefore, the average % of life years lost for individuals with disease onset during the study period (<42) is likely to be larger

than for those obtained from the NHS Executive (1996) report which included data from many individuals with disease onset after the age of 42.

A third study limitation relates to the indirect estimation of the lifetime decision endpoints associated with maternal depression. It was assumed that development measures identified in chapter three at age eleven were equivalent to measures in this chapter obtained at age ten. The difference in ages is unlikely to be problematic given the results of the growth curve model 1 in chapter three which identified consistent incremental effects of postnatal depression for children across ages 3-11. Meanwhile, the assumption of measurement equivalency between the SDQ and the RBS might be appropriate as: the SDQ was designed based on the RBS, both have highly correlated scores (r=0.88) (Goodman, 1997), and RBS scores were transformed to the same scale as SDQ scores for this research. A thorough discussion of the appropriateness of the methods applied to estimate the lifetime decision endpoints for children exposed to postnatal depression is included in chapter seven.

Fourth, while this study identified a strong association between child development and later life outcomes, a broad limitation of observational studies is the inability to establish causality. However, if these findings are considered alongside evidence in the current literature base the case for causality becomes much stronger. The study design was informed through lifespan developmental theory where lifetime outcomes are considered to be the direct consequence of childhood abilities (Overton, 2003), (Halfon and Hochstein, 2002), (Lerner and Busch-Rossnagel, 2013), and whilst sparse, randomised evidence is available to support this theory (Currie, 2001). The limitation of the causality assumption in the overall estimation of lifetime effects is discussed in more detail in chapter seven.

Finally, a considerable number of participants dropped out of the 1970 BCS. Drop out can lead to attrition bias if data is not missing completely at random (MCAR) – MCAR assumes the probability of participants leaving the study is not correlated with any characteristics of cohort members. The results of a logistic regression analysis by Mostafa and Wiggins (2014) suggest the MCAR assumption is unlikely to hold in the 1970 BCS. Individual characteristics including being male, unmarried, and having parents of low social class with fewer years of education were found to significantly increase the probability of non-response. Based on the associations identified in this study, individuals with characteristics fitting this profile might be expected to have relatively worse lifetime crime, education and health outcomes.

Even if it is assumed that missing observations have relatively worse lifetime outcomes, it is difficult to speculate on the direction of any potential attrition bias. If associations follow those identified in this study, then it would seem more likely that the missing data also included individuals with relatively worse development outcomes. In this instance, effect sizes might be

underestimated if individuals with the worst development *and* lifetime outcomes are those with the highest probability of non-response. Conversely, the observed effect sizes could be overestimated as there are feasible reasons why missing data might occur more frequently in individuals with relatively better development outcomes. For example, it may be that individuals with better cognitive development have higher wage expectations making them less likely to respond to questionnaires when their earnings do not meet these expectations. Future research could explore the problem of attrition bias by generating complete case databases using imputation methods, as demonstrated with the application of multiple imputation by chained equations (MICE) in chapter three.

5.5.5 Usefulness of Predictions for Decision Makers & Future Research

Bearing in mind the above limitations, the model predictions from this study might be questioned in terms of their usefulness to inform decision makers. Here, the commonly cited phrase by George Box (1976), that "all models are wrong, but some are useful", is particularly apt. The empirical models in this analysis have enabled estimates to be made regarding the effects of child development on later life outcomes based on logical simplifications of the real world.

The question of usefulness is to ask whether a model estimate is better than no estimate at all. NICE (2018) guidelines on maternal depression screening recognise the potential impact of postnatal depression on child development but terms these outcomes as "intangible". The impact of including/excluding lifetime outcomes of maternal depression on child development will be investigated in chapter six when assessing the cost-effectiveness of screening for postnatal depression. The usefulness of predictions generated in this chapter will be revisited in chapter seven.

5.5.6 Conclusions

The empirical analysis in this study identified strong associations between cognitive and socioemotional development and several lifetime outcomes that are likely to be important for healthcare decision making in the UK. Predictions from the models can be used to estimate lifetime effects of childhood health technologies and could be informative for decision makers assessing cost-effectiveness.

The research in this chapter estimated the lifetime effects of children exposed to postnatal depression, suggesting a substantial population burden occurring predominantly in the education sector through effects to economic productivity. The lifetime effects estimates

obtained from this chapter are used in the final part of the applied research in this thesis (Chapter 6) as parameters in a decision model evaluating the cost-effectiveness of screening for postnatal depression.

Future research could address some of the study limitations by using a more sophisticated path model. Researchers might also attempt to replicate findings in different longitudinal cohorts, which could increase the external validity of these results, reduce uncertainty around estimates and increase their usefulness when informing decisions.

5.6 Acknowledgements

Data from the British Cohort Study 1970 (UoL, 2016) were supplied through the UK Data service archive at the University of Essex. All results and interpretations of this analysis are the view of the author.

Chapter 6: An Economic Evaluation Assessing the Cost-Effectiveness of Screening for Postnatal Depression

6.1 Summary

The overarching aims of this thesis focus on developing a methodology for estimating the lifetime effects of early childhood circumstances and determining whether lifetime estimates can be used to inform cost-effectiveness decisions in UK economic evaluations. This chapter addresses these research aims presenting the results of an economic evaluation assessing the cost-effectiveness of competing screening strategies for postnatal depression. The parameters in this analysis are informed by the research in the previous chapters which estimated the lifetime incremental effects of childhood exposure to postnatal depression symptoms on health and economic decision endpoints. This chapter investigates the overall usefulness and influence of the lifetime estimates for decision-makers by identifying the cost-effective screening strategy for scenarios which either include or exclude the child parameters. The analysis also explores the importance of the philosophical framework applied in the economic evaluation of child health technologies by considering the screening strategy for postnatal depression from a health centric and cross-sectoral decision maker's perspective. As well as informing the overarching aims of this thesis, the applied research in this chapter can be used to update current policy recommendations regarding the appropriate screening strategy for postnatal depression in the UK.

6.2 Introduction

6.2.1 Background to Screening

Screening is a clinical method used to identify a sub group of a population at high risk of having a disease/disorder. Following a positive screen this sub group will usually proceed to further health technology interventions including diagnostic tests and/or the offer of an appropriate treatment (Gilbert et al., 2001). Therefore, screening allows for the timely identification and treatment of conditions which may not yet display clinical symptoms. Screening can be particularly beneficial for diseases with high population prevalence that may be left undiagnosed (Goldberg, 2014).

Screening is seldom perfect and typically leads to the identification of four distinct groups. Two of these are correctly classified: true positives who have the condition and have a positive screen and true negatives who do not have the condition and have a negative screen. The other two

groups are the misclassified false positives who do not have the condition but achieve a positive screen, and false negatives who have the condition but are not identified by the screen. These classifications are often summarised in terms of the screen's *sensitivity,* which is the ability of the test to detect the condition (true positives/ (true positives + false negatives)), and *specificity*, the ability of the test to identify those without the condition (true negatives/ (true negatives + false positives)) (Greenhalgh, 1997), (Lalkhen and McCluskey, 2008).

6.2.2 Screening for Postnatal Depression

There is some debate in the literature regarding the benefit of screening for postnatal depression (Paulden et al., 2009). On one hand screening fulfils many of the twenty criteria for viable population screening set out by the UK National Screening Committee (UK NCS, 2015). For example, postnatal depression is an important health condition with a high population prevalence (Sharp et al., 2010), has available and relatively non-invasive screening instruments whose sensitivity and specificity are established in randomised trials (Hewitt et al., 2009), and has cost-effective treatments available in the form of psychological and pharmacological therapies (NICE, 2018).

However, a study by Paulden et al. (2009) and the associated Health Technology Assessment (HTA) submission by Hewitt et al. (2009) did not find screening for postnatal depression to be cost-effective measured against cost-effectiveness thresholds commonly used in the NHS. Hewitt et al. (2009) suggests this result is primarily due to the cost of treating false positive cases. This evidence may have contributed to the UK National Screening Committee's (UK NCS, 2011) decision not to recommend population screening for postnatal depression.

More recently, NICE (2018) have investigated the cost-effectiveness of a dual phase screening strategy which reduces the number of costly false positive diagnoses. The suggested strategy implements a low costing screen with very high sensitivity (proportion of correct true positives) but low specificity, (proportion of correct true negatives) followed by a more thorough screen with higher specificity – the second screen being limited to positive cases identified through the initial screen (true positives and false positives). When compared to case identification through standard care, NICE (2018), find the dual strategy to be cost-effective when using the Whooley questions as the initial screen and either the Patient Health Questionnaire-9 (PHQ-9) or Edinburgh Postnatal Depression Scale (EPDS) as the second screen.

While the dual screening strategy is identified as cost-effective, current NICE (2018) guidelines do not *require* the formal administration of a population screen but suggest GPs *consider* asking these questions during a woman's first contact with primary care after child birth.

6.2.3 Research Objectives

To recap, the overarching aims of the research in this thesis are: to develop and describe a methodology for indirectly estimating the lifetime effects of early childhood circumstances by linking results across two empirical analyses; and to determine whether lifetime estimates influence the cost-effectiveness results in an applied, UK based, economic evaluation.

The primary objective of this chapter is to address the second research aim by presenting the results of an economic evaluation assessing the cost-effectiveness of screening for postnatal depression with and without the inclusion of lifetime effects for children exposed to symptoms of postnatal depression. The influence of lifetime effects is investigated for two separate decision perspectives as was suggested in chapter two. Firstly, an economic evaluation is conducted from the perspective of a health centric decision maker, limiting the decision endpoints to health benefits valued as QALYs and healthcare costs incurred directly by the NHS and PSS. Secondly, the perspective is extended to that of a cross-sectoral decision maker, adopting a perspective which extends the decision endpoints to include monetized costs and benefits falling on sectors outside of health.

In addition to addressing an overarching thesis aim, a second objective of the research in this chapter is to provide evidence which could be used to inform a UK health policy decision. Current recommendations for postnatal depression screening in the UK are informed through an economic evaluation conducted by NICE (2018). The analysis in this chapter extends the NICE (2018) model by addressing two limitations: firstly NICE (2018) did not formally account for the effects of postnatal depression in children; secondly, NICE (2018) did not assess cost-effectiveness across a full range of threshold values shown to influence the specificity and sensitivity of the EPDS/PHQ-9 screening instruments (Hewitt et al., 2009). The analysis in this chapter assesses the cost-effectiveness of screening for postnatal depression across all cut-off points for which data is available.

The final objective of this chapter is to determine whether future research is worthwhile and to establish research priorities if this is the case. Alongside policy implementation recommendations, economic evaluations should assess whether it is cost-effective to conduct future research (Drummond et al., 2015). With respect to the aims of the thesis, the analysis establishes the potential cost-effectiveness of future research obtaining more precise estimates of lifetime effects in children exposed to postnatal depression.

The above objectives were addressed through the following research questions:

1. What is the cost-effective screening strategy for postnatal depression when considering maternal outcomes only and assuming a healthcare decision maker's perspective?

2. What is the cost-effective screening strategy for postnatal depression when considering maternal and child outcomes and assuming a healthcare decision maker's perspective?
3. What is the cost-effective screening strategy for postnatal depression when considering maternal and child outcomes and assuming a cross-sectoral decision maker's perspective?
4. Is it cost-effective to obtain future evidence and should this evidence include further research regarding the estimation of lifetime outcomes associated with children who are exposed to postnatal depression?

6.3 Methods

6.3.1 Decision Analytic Model

The economic model in this analysis predominantly follows the same structure, adopts the same underlying assumptions and is informed by the same model parameters as the previously mentioned NICE (2018) model. Any departure from NICE (2018) methodology is distinguished using *italics*.

6.3.1.1 General Model Structure

Summarising research by Sculpher et al. (2006), the conclusions from chapter two suggest decision analytic models (DAMs) are the most appropriate vehicle for health economic evaluation. One of the most common types of DAM is the decision tree which was the model type selected by NICE (2018). Decision trees map all potential decisions (i.e. treatment strategies) and subsequent patient outcomes on different branches within the tree. Decision endpoints and probabilities are assigned to each patient outcome and mean costs and benefits are calculated for each strategy by simulating a hypothetical population through the model, multiplying the proportion of the population achieving each patient outcome by the assigned decision endpoints and summing across all the relevant branches (Brennan et al., 2006).

The decision tree in this analysis compared four formal case identification (screening) strategies for postnatal depression against the comparator of case identification through standard GP care. It was assumed that all screening strategies would be administered six weeks postnatally – this was selected as the optimal time for screening in analyses by NICE (2018) and Hewitt et al. (2009) as it is the period of peak incidence for postnatal depression where new mothers continue to have regular interaction with primary care services through GP appointments and home health visits.

Following screening, mothers could be grouped into four possible treatment pathways up to the model's endpoint at 52 weeks. The model included the (hypothetical) population of all postnatal women in the UK who are eligible to use the NHS. As informed by NICE (2018), postnatal depression was categorised as women with sub threshold, minor and major depression symptoms, occurring at a prevalence of 8.7%.

6.3.1.2 Screening Pathways

The first screening strategy was the Edinburgh Postnatal Depression Scale (EPDS), a ten- item instrument developed specifically to detect cases of depression in women during the antenatal and postnatal period (Cox et al., 1987). The ten items in the EPDS relate to symptoms of depression (e.g. misery, anxiousness, irritability and panic) and each is scored based on the frequency of symptom occurrence from zero (never) to three (most of the time/a lot). A total score is generated by summing scores across all items and ranges from zero to thirty. Upon conception Cox et al. (1987) suggested that probable cases of depression could be identified using a threshold of 12/13 with possible cases using a threshold of 9/10.

As identified in a meta-analysis by Paulden et al. (2009) and Hewitt et al. (2009), the sensitivity and specificity of the EPDS is dependent on the selected threshold. The NICE (2018) model did not assess cost-effectiveness across the full range of threshold values shown to influence the specificity and sensitivity of the EPDS/PHQ-9 screening instruments, instead investigating the EPDS as a single strategy using a combined threshold of 9 or 10. It was therefore limited in scope.

Consequently, *this analysis* extended the NICE (2018) screening model by assessing a range of thresholds for the EPDS (from 7-16) as competing strategies. The sensitivity and specificity of each EPDS threshold was obtained directly from Paulden et al. (2009) and is summarised in Table 6.1.

The second screening strategy investigated administration of the Patient Health Questionnaire 9 (PHQ-9). The PHQ-9 contains nine items for each of the DSM-IV criteria for depression, which are scored from zero ("not at all") to three ("nearly every day"), a total score ranging from 0-27 is obtained by summing scores for each item. Whilst the PHQ-9 is most commonly used as a tool to for identifying depression in the general population, it has reasonable sensitivity (0.75) and specificity (0.88) in postnatal women at threshold value of 9/10 (NICE, 2018). The PHQ-9 strategy was limited to this single threshold as no studies have identified the sensitivity and specificity of the PHQ-9 in postnatal women at a range of threshold values (NICE, 2018).

Table 6.1: Sensitivity and Specificity of Case Identification Strategies

Strategy	Mean	SE	Probabilistic Distribution	Source
Sensitivities				
EPDS (threshold=7)	0.9117		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=8)	0.9120		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=9)	0.8528		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=10)	0.8170		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=11)	0.7221		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=12)	0.6805		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=13)	0.6619		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=14)	0.5331		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=15)	0.3910		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=16)	0.3127		Log Normal, Cholesky*	(Paulden et al., 2009)
Whooley Questions	0.9500	0.04	Beta (α=34, β=2)	(Bosanquet et al., 2015)
PHQ-9 (threshold=9/10)	0.7500	0.06	Beta (α=44, β=15)	(NICE, 2018)
Standard Care	0.5100	0.05	Beta (α=52, β=50)	(NICE, 2018)
Specificities				
EPDS (threshold=7)	0.6699		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=8)	0.7454		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=9)	0.8216		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=10)	0.8626		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=11)	0.9110		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=12)	0.9306		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=13)	0.9271		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=14)	0.9572		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=15)	0.9776		Log Normal, Cholesky*	(Paulden et al., 2009)
EPDS (threshold=16)	0.9893		Log Normal, Cholesky*	(Paulden et al., 2009)
Whooley Questions	0.6500	0.05	Beta (α=69, β=37)	(Bosanquet et al., 2015)
PHQ-9 (threshold=9/10)	0.8800	0.02	Beta (α=396, β=54)	(NICE, 2018)
Standard Care	0.8100	0.03	Beta (α=102, β=23)	(NICE, 2018)

Notes: No SE is reported for EPDS by Paulden et al. (2009), variances on the Log Odds scale are available in the original report.
*Cholesky matrix was obtained through decomposition of variance-covariance matrices which were obtained through correspondence with authors.

The third and fourth case identification strategy consisted of two stages: an initial screen using the Whooley questions, followed by a second screen with either the EPDS or PHQ-9 for the subgroup of mothers who were categorised as positive cases by the first screen. The Whooley questions consist of two items asking mothers if they have (i) often been bothered by feeling down, depressed or hopeless and (ii) had little interest or pleasure in doing things. Affirmative responses to either of the questions result in a positive test (Bosanquet et al., 2015). The Whooley questions have a very high sensitivity (0.95) but low specificity (0.65) for depression in the general population (Bosanquet et al., 2015) and given the short administration time the combination of Whooley questions followed by a more specific instrument was considered as a potentially cost-effective strategy by NICE (2018).

The NICE (2018) model assumed the Whooley questions had a sensitivity of 1.00 based on the results from two studies using a population of postnatal women (Gjerdingen et al., 2009),

(Mann et al., 2012). *This analysis* updated these parameters using evidence by Bosanquet et al. (2015) who identified the specificity and sensitivity of the Whooley questions for detecting depression in a general population in a ten study meta-analysis including studies for postnatal populations by Gjerdingen et al. (2009) and Mann et al. (2012). It was decided that evidence from the eight additional studies in the general population would be relevant for the postnatal population and provide more precise estimation of sensitivity/specificity. As previously, *this analysis* investigated a variety of EPDS threshold values (ranging from 7-16) during the second screen for positive cases identified by the Whooley questions.

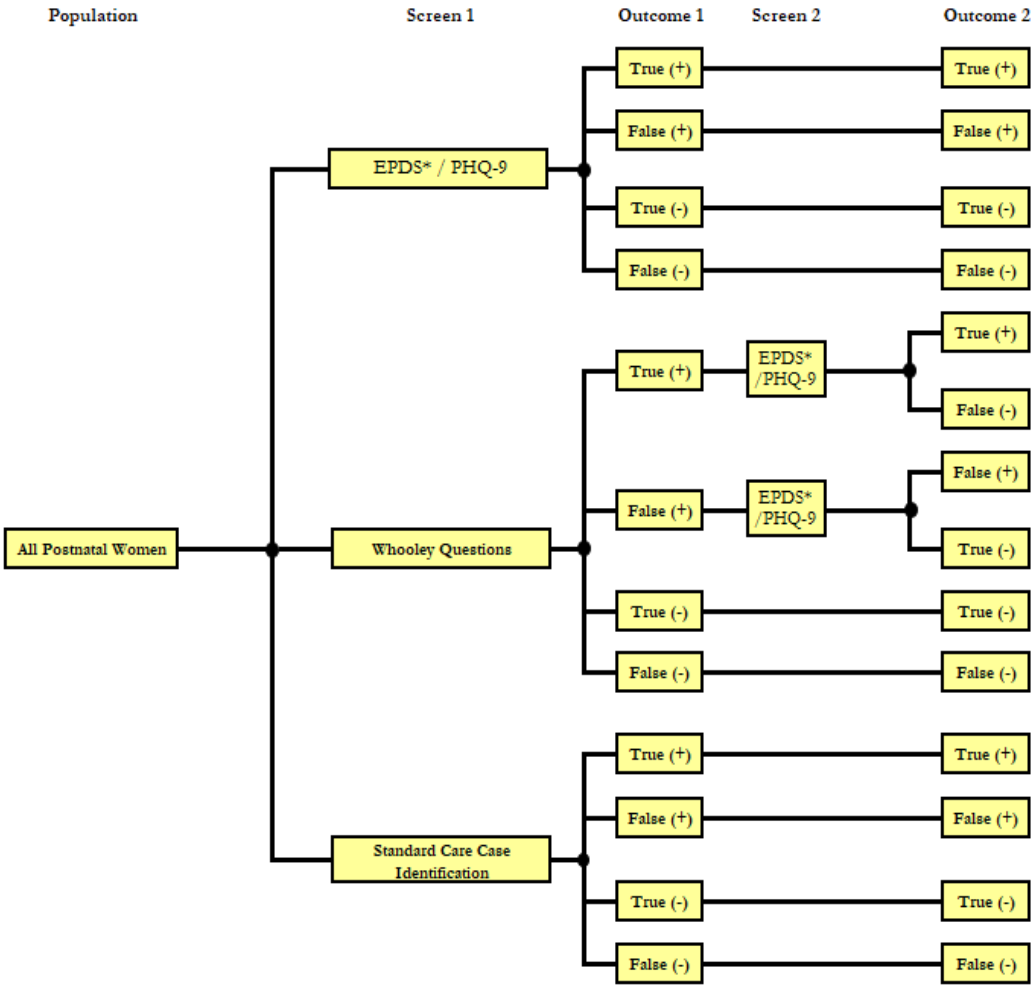


Figure 6.1*:* Illustrates the diagnostic component of the decision tree.
*='Ten different strategies were investigated for different thresholds for the EPDS ranging from 7-16.

The final strategy was the comparator, standard care case identification which assumed that a proportion of postnatally depressed women would be identified as depressed by a GP through

contact with the health service during the postnatal period. Standard case care identification was assumed to occur with a sensitivity of 0.51 and specificity of 0.81 as identified by NICE (2018).

Each case identification strategy established four distinct groups which followed diverging treatment components: "true positives" were depressed mothers who correctly identified; "false positives" were non-depressed mothers who had a positive screen; "true negatives" were non-depressed women who had a negative screen; while "false negatives" were depressed mothers who were not identified through screening. The strategies assessed in this analysis and the associated groups identified through screening are depicted in Figure 6.1.

6.3.1.3 True Positive & False Positive Treatment Component Assumptions

Both the true positive and false positive mothers were assumed to receive treatment for depression and therefore followed the same treatment components. It was expected that allocated treatments would depend on the severity of depression. Using prevalence rates reported by (NICE, 2018), it was assumed that 72% of women suffered from mild-moderate symptoms whilst 28% of women suffered from moderate-severe symptoms.

In cases of mild-moderate depression, women were assumed to be offered facilitated guided self-help. This is a treatment which teaches individuals to personally tackle symptoms of depression using cognitive behavioural strategies, delivered through self-help manuals with assistance being provided by therapists, computer software, and web applications (NICE, 2018). Women were expected to complete seven sessions over a period of ten and a half weeks, this being the mean reported in NICE (2018) guidelines. In addition, mild-moderately depressed women were also expected to attend three GP appointments during this period.

Cases of moderate-severe depression were expected to be offered either intensive psychological therapy (72%), or pharmacological treatment (28%). It was assumed that psychological therapy would consist of 16 sessions of cognitive behavioural therapy conducted over a three-month period and be delivered by a trained therapist (NICE, 2018). Women receiving pharmacological treatments were assumed to receive sertraline (or similar antidepressants) for eight weeks of intensive treatment followed by six months of maintenance therapy. Fifteen percent (15%) of mothers receiving pharmacological therapy were assumed to have two specialist psychiatric appointments while the remainder (85%) were managed in primary care with an expected mean of four GP appointments (NICE, 2018).

At 52 weeks, true positive mothers were categorised as depressed/not depressed depending on whether they responded or failed to respond to treatments. The absolute risk of non-response for facilitated self-help (0.4891), intensive psychological therapy (0.3120) and pharmacological therapy (0.3120) were not directly reported by NICE (2018) and were therefore derived by

multiplying the reported relative risks with the reported absolute risk of no improvement in standard care (mild-moderate depression=0.65, moderate-severe depression=0.67) (NICE, 2018). As all false positive mothers were wrongly identified as depressed, they were assumed to follow the "response" pathway and were categorised as non-depressed at the model endpoint. The true positive/false positive treatment components are illustrated in Figure 6.2.

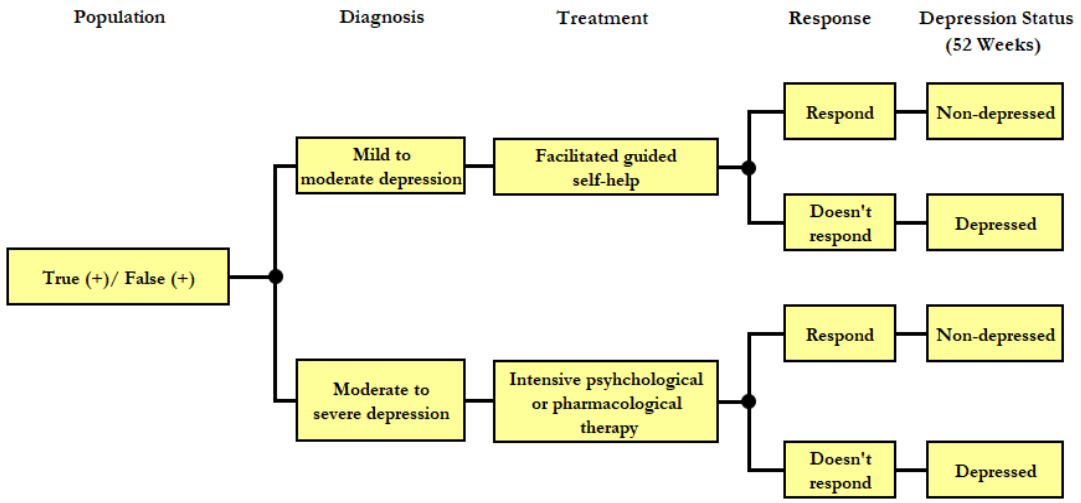


Figure 6.2*:* Depicts the true positive and false positive treatment components of the decision tree.

6.3.1.4 True Negative Treatment Component Assumptions

The true negative population consisted of mothers who were correctly identified as not depressed and, as such, were assumed to receive no treatment for depression throughout the duration of the decision model. This true negative treatment component is illustrated in Figure 6.3.

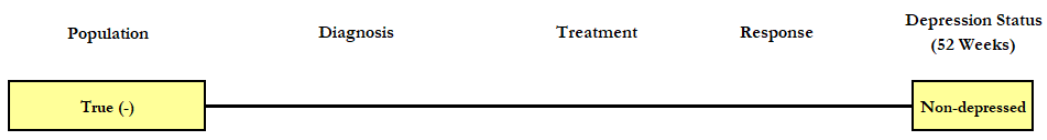


Figure 6.3: Illustrates the true negative treatment component of the decision tree

6.3.1.5 False Negative Treatment Component Assumptions

Depressed mothers who were wrongly identified as not depressed followed the false negative treatment component as is illustrated in Figure 6.4. It was assumed that 33% of these mothers would spontaneously recover from postnatal depression three months following child birth without receiving any formal treatment, and that they would remain in the non-depressed disease state throughout the remainder of the model (NICE, 2018).

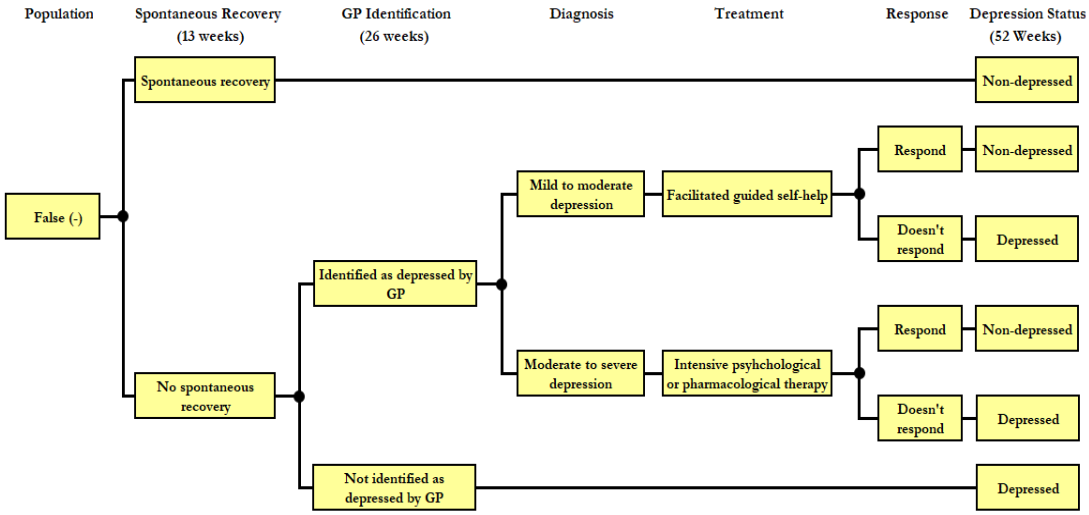


Figure 6.4: The false negative treatment component of the decision tree.

The false negative population who did not spontaneously recover were either identified (8%) or not identified (92%) as depressed by a GP at six months postpartum. Those who were not identified by a GP received no treatment and were categorised as depressed at 52 weeks. Mothers who were identified as depressed by a GP were assumed to immediately begin treatment, following the same pathway as true positive/false positives (NICE, 2018). All probabilities used within the economic model are described in Table 6.2.

Table 6.2: Probability Parameters

Parameter	Mean	SE	Probabilistic Distribution	Source
Postnatal depression prevalence	0.087	0.004	Normal	(NICE, 2018)
Mild-moderate symptoms |depressed	0.725	0.030	Normal	(NICE, 2018)
Moderate-severe symptoms |depressed	0.275	0.030	Normal	(NICE, 2018)
Absolute risk of no response Standard care (mild-moderate symptoms)	0.67	0.020	Normal	(NICE, 2018)
Absolute risk of no response in standard care (moderate-severe symptoms)	0.65	0.020	Normal	(NICE, 2018)
Relative risk of no improvement facilitated self-help (vs. standard care)	0.73	0.100	Log Normal	(NICE, 2018)
Relative risk of no improvement psychological therapy (vs. standard care)	0.48	0.050	Log Normal	(NICE, 2018)
Relative risk of no improvement pharmacological therapy (vs. standard care)	0.48	0.050	Log Normal	(NICE, 2018)
Absolute rate of spontaneous recovery at 13 weeks	0.33	0.004	Normal	(NICE, 2018)
Absolute rate of GP identification at 26 weeks	0.08	0.040	Beta (α=4, β=50)*	(NICE, 2018)

Notes: *Probabilistic values obtained directly from Kessler et al. (2002), a study used to inform the NICE (2018) model.

6.3.2 Model Parameters

6.3.2.1 Maternal Healthcare Cost Assumptions

All maternal healthcare costs were obtained from the original NICE (2018) report and uprated to 2018 prices using the consumer price index (*ONS*, 2018b). Overall treatment costs (including additional care) equalled £386.87 for facilitated guided self-help, £231.08 for pharmacological therapy and £1710.14 for intensive psychological therapy. It was assumed that false positive cases would discontinue treatment before its conclusion, incurring 20% of the total treatment costs.

An additional cost of £8.82 for health and social care was applied for each week mothers were in the depressed disease state. This additional cost excluded the first six weeks following child birth, as it was assumed that both depressed and non-depressed mothers would have similarly frequent contact with the health services meaning net costs over this period would equal zero.

Costs of screening were estimated based on the hourly (uprated) cost for a GP consultation (£250.19). The EPDS/PHQ-9 screen was expected to take 15 minutes to complete at a cost of £62.55, while the Whooley questions were assumed to take 1 minute to administer costing £4.17. The duration of standard care case identification was set equal to the mean GP appointment time (11.7 minutes) and therefore cost £48.79. A further cost of £52.67 for an hourly appointment with a health visitor was applied to all positive cases that were identified through screening. A full description of all costing categories is provided in NICE (2018).

6.3.2.2 Maternal Quality Adjusted Life Years Assumptions

All utility scores reported by NICE (2018) were obtained from a trial conducted by Sapin et al. (2004) and were identified using the EQ-5D questionnaires in a general UK population. Postnatally depressed mothers were assumed to incur the same utility decrement as depressed individuals in the general population and thus assigned utility scores of 0.74 for mild-moderate symptoms and 0.44 for moderate-severe symptoms. Non-depressed mothers were assigned a health-related quality of life score of 0.86 which was associated with the remission health state in Sapin et al. (2004).

The NICE (2018) analysis did not report the QALYs for each maternal health state directly, instead stating QALYs as being calculated using the "area under the curve" method. Therefore, this analysis derived QALYs by applying the trapezium rule[9] to estimate the integral across the appropriate HRQoL utility scores. As the model was conducted over 52 weeks, QALYs were equivalent to HRQoL utilities for mothers who remained solely in one disease state (i.e. never depressed/depressed and no treatment response).

Mothers who responded to treatment did not have the same HRQoL scores throughout the model as they transitioned from the depressed to non-depressed disease states. The NICE (2018) methodology showed this transition occurring as a linear improvement in HRQoL over the duration of treatment. An example of the linear improvement in HRQoL is depicted in Figure 6.5 for a mother with moderate-severe depression who responds to intensive psychological therapy. Several QALYs values were derived in this analysis (using the trapezium rule) for maternal responders who differed in terms of on initial depression severity, the date of treatment initiation, and the total duration of treatment.

[9] For a full description of the trapezium rule see chapter five, section 5.3.3.6

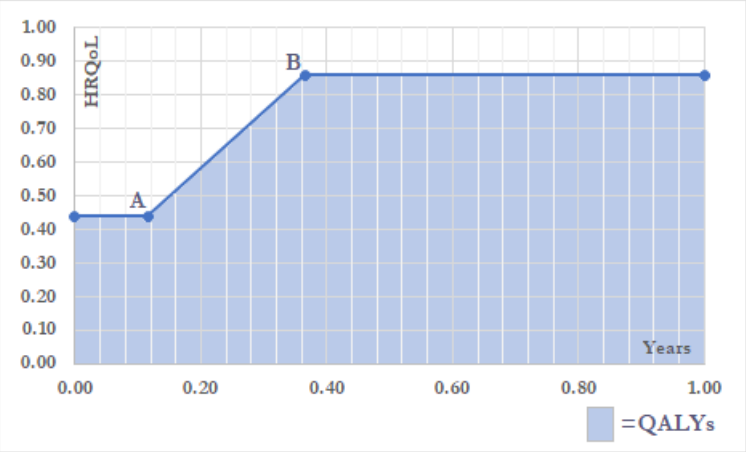


Figure 6.5: Illustrates the HRQoL and associated QALYs for mothers who initially suffered from symptoms of moderate-severe depression and who responded to three months of intensive psychological therapy during the model. A=treatment initiation at 6 weeks, B=treatment conclusion at 19 weeks.

The final maternal QALYs variable was assigned to false positive mothers who were assumed to be non-depressed but incur a 2% reduction in utility due to a wrong diagnosis. This assumption was obtained from the NICE (2018) model which was informed through NICE Guideline Development Group (GDG) expert opinion. Discounting was not applied to maternal healthcare costs or QALYs as the duration of the model was over a single year. A summary of all costs and utility parameters is provided in Table 6.3

Table 6.3: Maternal Cost and Utility Parameters

Parameter	Mean	SE	Probability Distribution	Source
Maternal Costs				
Facilitated guided self-help	£386.87	N/A	Fixed	(NICE, 2018)
Intensive psychological therapy	£1710.14	N/A	Fixed	(NICE, 2018)
Pharmacological therapy	£231.08	N/A	Fixed	(NICE, 2018)
Weekly health & social care[1]	£8.82	N/A	Fixed	(NICE, 2018)
Treatment costs for false positives[2]	20.0%	N/A	Fixed	(NICE, 2018)
EPDS/PHQ-9 screen	£62.55	N/A	Fixed	(NICE, 2018)
Whooley screen	£4.17	N/A	Fixed	(NICE, 2018)
Standard care case identification	£48.79	N/A	Fixed	(NICE, 2018)
Maternal Utilities				
No depression symptoms	0.86	0.13	Beta (α=215, β=35)	(NICE, 2018)
Mild-moderate depression symptoms	0.74	0.19	Beta (α=133, β=47)	(NICE, 2018)
Moderate-severe depression symptoms	0.44	0.27	Beta (α=31, β=39)	(NICE, 2018)
Utility decrement for false positives[3]	2.00%	N/A	Beta (α=2, β=98)	(NICE, 2018)

Notes: 1: Weekly health and social care costs for depressed mothers only
2: Treatment costs are a percentage of the total treatment costs for true positives
3: Percentage utility decrement from utility for "no depression symptoms"

6.3.2.3 Children's Lifetime Outcomes Assumptions

This analysis departs from the NICE (2018) methodology by including children's lifetime outcomes in the decision tree. The theoretical literature indicates children are affected by postnatal depression through the detrimental influence of symptoms on the maternal-child attachment relationship (Bowlby, 1978), (Cummings and Davies, 1994), (Sanger et al., 2015). In support of this assertion, a meta-analysis by Cuijpers et al. (2015) identifies significant benefits from psychological treatments for postnatal depression on maternal depression outcomes, maternal-child attachment outcomes and on early child development outcomes. This analysis assigned incremental lifetime outcomes by classifying children as being "exposed" or "not-exposed" to postnatal depression symptoms.

Children were assumed to be exposed to postnatal depression symptoms if their mothers were in the "depressed" disease state at 52 weeks. The primary analysis assumes that children were not exposed to symptoms of postnatal depression if their mothers were never depressed or initially depressed but responded to treatment (i.e. successful treatment assumed to prevent the effects of postnatal depression). This was assumed as chapter three identified postnatal depression at 9- months after child birth, resulting in a control group of "non-depressed" mothers which also contained a group of responders within it (mothers who had symptoms shortly after birth but did not have symptoms by the 9-month measurement point) This assumption is investigated in the deterministic sensitivity analysis. Figure 6.6 illustrates the assignment of exposure status groups for the true positive/false positive treatment component.

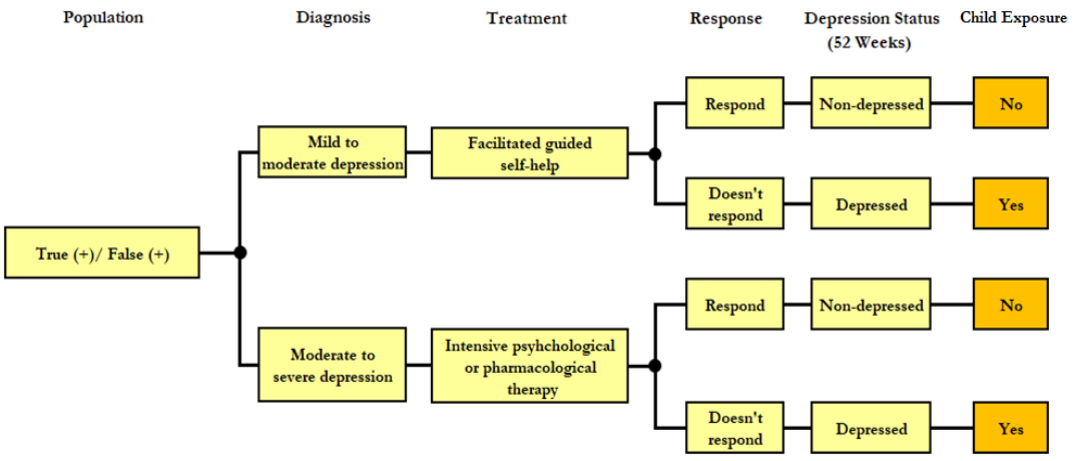


Figure 6.6: True positive/false positive treatment components including effects to children.

Table 6.4: Children's Lifetime Costs and Utilities

Parameter	Mean	SE	Probability Distribution	Source
Intermediate Development Effects[1]				
PND on GCA	-0.27	0.63	Normal, Cholesky	Chapter 3
PND on SDQ	2.07	0.28	Normal, Cholesky	Chapter 3
Development to Lifetime Outcomes[2]				
QALYs				
RBS	-0.004	0.003	Normal, Cholesky	Appendix 5.4
GCA	0.003	0.001	Normal, Cholesky	Appendix 5.4
Healthcare costs				
RBS	£58.76	£22.53	Normal, Cholesky	Appendix 5.4
GCA	-£15.81	£6.71	Normal, Cholesky	Appendix 5.4
Returns to Education				
RBS	-£625.13	£2,063.10	Normal, Cholesky	Appendix 5.4
GCA	£4,524.81	£701.63	Normal, Cholesky	Appendix 5.4
PND on Lifetime Outcomes[3]				
QALYS	-0.01			Derived
Healthcare Costs	£125.61			Derived
Returns in Education	-£2,512.52			Derived

Notes: 1: Incremental effects of exposure to postnatal depression symptoms at 9-months vs. non-exposure on child development outcomes obtained from Chapter 3, growth curve model 1.
2: Marginal effects per unit increase in child development measures on lifetime outcomes, obtained from mathematical models in Chapter 5 (Appendix 5.4).
3: Incremental effect of exposure to postnatal depression symptoms at 9-months vs non-exposure on lifetime outcomes derived by multiplying intermediate development effect sizes (Chapter 3) by marginal effect coefficients (Appendix 5.4).

Children exposed to symptoms of postnatal depression were assigned estimates of incremental lifetime QALYs, healthcare costs and returns to education as previously identified in chapter five. As the estimates were incremental, non-exposed children were assigned lifetime outcomes equal to zero. The primary analysis assigned incremental lifetime outcomes discounted at 3.5%, equalling -0.01 QALYs, £125.61 healthcare costs, and -£2,512.52 returns to education. Costs to the criminal justice sector were not included in the decision model as the added difficulty of incorporating costs across an additional sector was not considered worthwhile given the insubstantial effect sizes identified in chapter five. It was assumed that lifetime estimates obtained in this thesis at nine-months would be applicable to exposure occurring at 12-months in this model. All child model parameters are reported in Table 6.4.

6.3.3 Analytical Perspectives

Two different perspectives were adopted in this research. The first two objectives were investigated by adopting a health centric decision maker's perspective which accounted for health benefits (QALYs), and healthcare costs incurred by the NHS and PSS. For the first objective, only maternal health outcomes were considered, whilst the second objective

expanded these outcomes to include both maternal and child health outcomes. The third objective was investigated by adopting the perspective of a cross-sectoral decision maker. This perspective accounted for child and maternal health outcomes but also included children's lifetime monetary returns in education as a decision endpoint.

6.3.4 Decision Rules

Economic evaluation should primarily inform policy through a deterministic analysis which accounts for the *mean* costs and *mean* benefits associated with each strategy (Claxton, 1999), (Drummond et al., 2015). Deterministic analyses establish cost-effectiveness by applying decision rules to appropriately rank competing strategies in terms of their incremental mean costs and benefits. Decision rules in economic evaluation typically require the calculation of the incremental cost-effectiveness ratio (ICER), this being incremental costs divided by incremental benefits for two competing strategies (Drummond et al., 2015), Equation 6.1a.

The first analysis, which adopted a health centric decision perspective, identified cost-effectiveness using dominance decision rules. Initially strongly dominated strategies were identified and excluded from the cost-effectiveness decision if they had lower associated mean QALYs and higher mean healthcare costs than a competing strategy. Next, extended dominance was calculated by ranking each remaining strategy in terms of mean costs and calculating ICERs between strategies of increasing costs. Extendedly dominated strategies were identified (and excluded) if their ICER (with a higher costing strategy) decreased (Drummond et al., 2015). For instance, if assessing the cost-effectiveness of strategies A, B and C, strategy B is extendedly dominated if the ICER between A and B is larger than the ICER between B and C, and the associated healthcare costs are strategy C > strategy B > strategy A.

The first analysis identified the cost-effective strategy by comparing the ICERs between all non-dominated/extendedly dominated strategies with a cost-effectiveness threshold (k). The cost-effectiveness threshold is the opportunity cost associated with health technologies (i.e. the resources which could have been spent elsewhere in the health sector) and the value of k equates to the monetary value placed by the health sector per each additional unit of health (i.e. QALY) (Woods et al., 2016). The cost-effective strategy was identified here as that with the highest mean benefit *and* an ICER (with the next non-dominated higher costing strategy) strictly less than the cost-effectiveness threshold. Deterministic cost-effectiveness results are reported in analysis for thresholds of k=£20,000 and k=£30,000. Whilst there is currently some debate in the literature as to the appropriate value for k (Claxton et al., 2013), UK economic evaluations usually follow NICE (2013) reference case guidelines and conduct cost-effectiveness analysis using these values for k (Woods et al., 2016).

Equation 6.1a: Decision Rules for a Health Decision Maker's Perspective

A strategy is considered cost-effective (when compared with standard care) if:

ICER: $[\Delta c_h/\Delta h] > k$

NHB: $[\Delta h - \Delta c_h/k] > 0$

Equation 6.1b: Decision Rules for a Cross-Sectoral Decision Maker's Perspective

A strategy is considered cost-effective (when compared with standard care) if:

NHB: $[\Delta h - \Delta c_h/k] - \Delta c_c/v > 0$

Or equivalently:

NCB: $v.[\Delta h - \Delta c_h/k] - \Delta c_c > 0$

Where:

Δh =	incremental QALYs
Δc_h =	incremental healthcare costs
Δc_c=	incremental education costs
k=	cost-effectiveness threshold (the additional costs that would displace 1 unit of health elsewhere in the healthcare system)
v=	The consumption value of health (the amount of consumption regarded as equivalent to 1 unit of health

Notes: Equations obtained from Claxton et al. (2010). Decision rules establish cost-effectiveness by comparing a strategy to a relevant comparator e.g. standard/usual care. ICERs is the incremental cost-effectiveness ratios; NHB is the net health benefit; NCB is the net consumption benefit.

The second analytical perspective required children's monetised returns to education to be considered within the decision rule. The literature review in chapter two identified a study by Claxton et al. (2010) suggesting a decision rule which can account for monetised outcomes outside of health when evaluating the cost-effectiveness of public health interventions. The recommended decision rule identifies incremental net health benefits for each treatment strategy versus standard care.

When applied to analyses which only include health related decision endpoints, the formula for Net Health Benefit (NHB), or equivalently net consumption benefit (NCB), is a simple rearrangement of ICER inequality, Equation 6.1a. Claxton et al. (2010) modify the NHB equation to account for both extra consumption costs and a consumption value for health (v), Equation 6.1b. As described in chapter two, it may be that £1 consumption in the health sector

is not equivalent to £1 consumption in other government sectors. The consumption value for health provides a mechanism to weight costs and benefits falling on the wider economy differently to the costs and benefits falling within the health sector. For the second analytical perspective NHBs/NCBs were calculated for all strategies vs. standard care using Equation 6.1b. The cost-effective strategy was identified as that with the largest overall net benefit.

As no current research establishing an appropriate value for the consumption value of health (v) was identified, results are presented across a range of potentially feasible values. This began with the assumption that v/k=1 meaning that £1.00 spent in the health sector was assumed equivalent to £1.00 spent in the education sector. The ratio of v/k was then decreased in increments to a minimum value of 0.25 where £1 health = £0.25 education and increased in increments to a maximum of value of 4.00 where £1 health= £4 education. Results at each v/k ratio are identified for thresholds of k=£20,000 and k=£30,000.

6.3.5 Evaluating Uncertainty

All economic evaluations using decision analytic models contain some form of uncertainty through methodological choice or through imprecision in the estimation of the model (Drummond et al., 2015). Common sources of uncertainty include: structural uncertainty, usually resulting from the assumptions made when building the model; and parameter uncertainty resulting from imprecision in empirical estimates used to inform parameter values i.e. confidence intervals around sample means from clinical trials/observation studies (Andronis et al., 2009), (Drummond et al., 2015). Therefore, it is important, and a requirement for submissions to NICE (2013), that appropriate sensitivity analyses be conducted to characterise this uncertainty.

6.3.5.1 Deterministic Sensitivity Analysis

One method of characterising uncertainty is through deterministic sensitivity analysis which assigns a specific input value to model parameters in exchange for the original parameter value. This type of sensitivity analysis is particularly useful for identifying the impact of structural uncertainties as it allows cost-effectiveness to be assessed in different scenarios by relaxing or adjusting the assumptions made when building the model (Drummond et al., 2015).

Deterministic sensitivity analysis was undertaken for several questionable assumptions in this analysis. Firstly, the assumption that all false positive mothers receive a 2% utility decrement from being wrongly diagnosed was based on GDG expert opinion and not on empirical evidence (NICE, 2018). Therefore, the first scenario investigated the case where false positive mothers achieved the same utility as true negative mothers (i.e. no utility reduction).

The second scenario investigated assumptions regarding the transmission of treatment benefit from mother to child. The primary analysis assumed that children were not exposed to symptoms if their mother recovered from postnatal depression during the model. In contrast, the sensitivity analysis assumed children were affected by depression symptoms proportionate to the duration their mother was in the depressed disease state. For instance, if a mother was in the depressed disease state for three months and was non-depressed for nine months, their child incurred 25% of the total healthcare cost and QALYs of a child whose mother was depressed for the full 52 weeks.

Thirdly, the lifetime effects of children exposed to symptoms of postnatal depression were indirectly estimated in this thesis by assuming exposure effects on development outcomes continued and remained consistent throughout the child's lifespan. This assumption was informed by the results of Growth Curve Model 1 in chapter three, which identified equivalent incremental effects on GCA and SDQ outcomes for all observations for children aged 3-11. Whilst incremental effects appeared to remain consistent within the study period, it is possible that the effects of postnatal depression changed and had either an increasing or decreasing impact on children beyond the study time horizon observed in chapter three. To investigate the influence of this assumption the cost-effective screening strategy was identified for scenarios which apply an (arbitrary) 25% reduction/increase to all children's lifetime outcomes.

Deterministic sensitivity analysis allows cost-effectiveness to be established for a range of sub-populations (Andronis et al., 2009). For postnatal depression there is evidence that prevalence rates vary according to demographic characteristics, for instance, first time mothers with low income are more likely to suffer from postnatal depression than the general population (Goyal et al., 2010). Representative values for low income populations were investigated in a fourth sensitivity analysis by increasing the prevalence rates of postnatal depression from 8.7% to 20.0%, which is towards the upper bounds reported by Goyal et al. (2010).

Finally, as described in chapter five, there is debate regarding the value of the discount rate applied to decision endpoints in economic evaluation. All maternal decision endpoints were unaffected by discounting as they occurred within the 52-week time horizon. In contrast, child outcomes were identified across the lifetime and the results of chapter five estimated substantially smaller lifetime effects for lower discount rates. This research reports all primary results with costs, QALYs and monetised returns to education discounted at 3.5%. In addition, a deterministic sensitivity analysis was conducted identifying the cost-effective strategy for a scenario where all costs and benefits are discounted at 1.5%, and for a scenario where unequal discount rates are applied with healthcare costs and monetised returns to education discounted at 3.5% and QALYs discounted at 1.5%.

6.3.5.2 Probabilistic Sensitivity Analysis

To evaluate how parameter uncertainty affected the overall decision, probabilistic sensitivity analysis (PSA) was performed. This type of sensitivity analysis applies probability distributions to model parameters and draws *n* random sample from each of these distributions. The cost-effective strategy is identified by calculating the Net Health Benefit for each strategy in *each* random draw. PSA is used to determine the probability of a correct decision being made by identifying the probability of cost-effectiveness associated with each strategy. These are obtained by summing the total number of times a strategy is cost-effective across all draws and dividing this by the total number of draws (*n*) (Briggs et al., 2002), (Drummond et al., 2015), (Fenwick et al., 2001).

To conduct PSA, appropriate distributions need to be assigned to each model parameter. As model parameters are averages (i.e. means), the distribution selected for PSA should represent the expected sampling distribution for the parameter of interest rather than the expected distribution of the population parameter (Briggs et al., 2002).[10] Because PSA is concerned with sample distributions, there is a relatively small range of candidate distributions (Briggs et al., 2002).

Briggs et al. (2002) recommend the normal distribution as an option for most model parameters in PSA. According to central limit theorem, sample distributions are normally distributed if sample sizes are large enough even if the population distributions of the parameter are not. However, normal distributions may not be suitable for probabilities and utility values which are bounded by zero and one as random draws could identify model parameter values outside of this feasible range. For bounded model parameters, Briggs et al. (2002) suggests researchers consider the gamma distribution. As a continuous form of the binomial distribution, the gamma distribution can be fit to various probability/utility data by altering the distributional parameters α (number of outcomes) and β (number of trials), thus producing random draws within the possible bounds of the model parameter.

Similarly, normal distributions might not be appropriate if assigned to relative risk parameters as these are probability ratios (i.e. probability of outcome in exposed group: probability of outcome in control group). As described by Briggs et al. (2002), and according to central limit theorem, sample distribution for relative risks are expected to be normally distributed following

[10] A population distribution is simply the distribution (or occurrence) of the parameter of interest (e.g. mean) across the whole population. As it is not usually possible to measure outcomes across the entire population, empirical studies *estimate* population parameters by conducting analyses on sub-sections, or samples, of the population. A sampling distribution is the distribution of all the different sample means obtained from many different samples of a population.

the application of a natural log transformation. Consequently, the log normal distribution might be the most appropriate selection of distribution for relative risk parameters in PSA.

The distributions selected for this analysis were informed by Paulden et al. (2009) who conducted PSA in an economic evaluation assessing the cost-effectiveness of screening for postnatal depression. All distributional assumptions made by Paulden et al. (2009) were confirmed to be consistent with the above recommendations by Briggs et al. (2002). The distributions applied to each parameter are described alongside the associated parameters previously in Tables 6.1-6.4. Fixed distributions (i.e. non-random) were applied to treatment costs as there is little uncertainty regarding the value of these variables (Briggs et al., 2002).

A final consideration during the assignment of distributions in PSA is the possibility of dependencies between model parameters (Claxton et al., 2005), (Naveršnik and Rojnik, 2012). For instance, the specificity of a screening strategy is likely to be dependent on the test's sensitivity. Assumptions of independence between test sensitivity and specificity could lead to improbable outcomes for random draws within PSA, e.g. individual draws might assign a test with very high specificity *and* sensitivity. PSA can be conducted with correlated random variables to account for potential dependencies between model parameters (Claxton et al., 2005), (Naveršnik and Rojnik, 2012). Correlated random variables are obtained for PSA by algebraically manipulating the covariance-variance matrix for dependent model parameters using a technique called Cholesky decomposition (Briggs et al., 2006), (Naveršnik and Rojnik, 2012). It is only usually possible to apply Cholesky decomposition in primary analyses as the variance-covariance matrix is rarely reported in secondary literature.

This analysis identified correlated random variables for: all EPDS screening parameters where the covariance-variance matrix was obtained from Paulden et al.(2009) through liaison with authors; and for all child parameters where the covariance-variance matrix between cognitive and socioemotional variables was obtained for the analyses conducted in chapters 3 and 5. It was not possible to identify this information for the Whooley screens, which were assigned as single random variables with a beta distribution.

This analysis conducted PSA using 10,000 random draws. Probabilities of cost-effectiveness were established for cost-effective thresholds ranging from £0 - £100,000 per QALY. Separate PSA results were obtained for each decision perspective. PSA identifies probabilities of cost-effectiveness for each strategy which are dependent on the cost-effectiveness threshold. This is because cost-effectiveness is calculated using the Net Health Benefit equation which includes the cost-effectiveness threshold within it (see equation 6.1a). The results of PSA are presented as cost-effectiveness acceptability curves (CEACs) and cost-effectiveness acceptability frontiers (CEAFs) (Barton et al., 2008), (Fenwick et al., 2001).

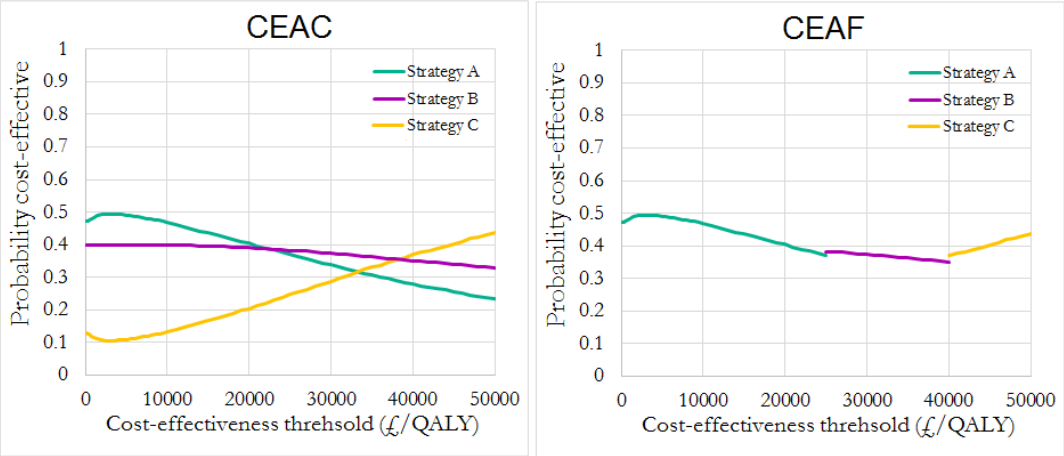


Figure 6.7: Illustrates an example cost-effectiveness acceptability curve (left) and cost-effectiveness acceptability frontier (right) for strategies A, B and C. CEACs display PSA results for all strategies across different cost-effectiveness thresholds, CEAFs display PSA results for the mean cost-effective strategy across cost-effectiveness thresholds. In this hypothetical example the mean cost-effective strategy is assumed to be: strategy A when k<£25,000; strategy B when £25,000<=k<£40,000; and strategy C when k>=£40,000.

Cost-effectiveness acceptability curves plot each strategy's probability of cost-effectiveness across the range of cost-effectiveness thresholds under investigation. Similarly, CEAFs plot the probability of cost-effectiveness across threshold ranges, but *only* plot this for a single strategy identified as cost-effective through deterministic analysis of *mean* model parameters. Therefore, assuming policy makers adopt the mean cost-effective strategy, the CEAF illustrates the associated probability of a correct decision being made across the feasible range of cost-effectiveness thresholds. An example illustration of CEACs and CEAFs is provided in Figure 6.7. Fenwick et al. (2001) recommend CEAFs when displaying PSA results as the mean cost-effective strategy is not always the strategy with the highest probability of being cost-effective.

6.3.5.3 Value of Information Analysis: EVPI

In addition to informing decisions about the cost-effectiveness of implementing a specific strategy, economic evaluation should also establish whether it is cost-effective to conduct further research (Drummond et al., 2015). This question can be addressed using the results of PSA to conduct a *value of information analysis* (Barton et al., 2008). The first figure estimated in a value of information analysis is typically the *expected value of perfect information* (EVPI). The EVPI calculates the cost of removing all the uncertainty from decision making and, therefore, represents the maximum value of information obtained through additional

research (Drummond et al., 2015). If the costs of future research are expected to exceed the EVPI then it should not to be commissioned as it will not be cost-effective.

Following the methodological guidance in Ades et al. (2004), Brennan et al. (2007) and Drummond et al. (2015), the EVPI was calculated in this analysis by applying the following procedures to the outputs of the PSA:

(1) The optimal decision given current information was identified as the strategy with the largest mean net benefit across *all* draws in the PSA. The largest mean net benefit given current information (NB_{ci}) entered the EVPI calculation.
(2) The expected net benefit of a decision taken when given perfect information was calculated by: firstly, identifying the maximum net benefit (for any strategy) within *individual* probabilistic draws and secondly calculating the mean of the individual maximum net benefits across *all* probabilistic draws. This calculated mean value was termed the mean maximum net benefit given perfect information (NB_{pi}).
(3) The final EVPI was calculated as NB_{pi} - NB_{ci}.

For further clarity Table 6.5 provides an example of an EVPI being calculated for a *hypothetical* cost-effectiveness analysis of strategies A, B and C.

The above procedure identifies the EVPI per individual. Healthcare policy decisions are primarily concerned with the appropriate allocation of healthcare at a population level. Therefore, it is most appropriate to present the EVPI results per population rather than per individual. The EVPI per population is obtained by multiplying the number of individuals expected to benefit from the strategy (patient population) by the value of the EVPI per individual (Ades et al., 2004), (Brennan et al., 2007).

Following methods by Paulden et al. (2009), this evaluation identified the patient population assuming screening would be offered to all new mothers from England and Wales and would continue to be offered for the next twenty years. ONS (2018a) data indicated that 685,320 mothers gave birth in England and Wales during 2018. The cumulative number of potential beneficiaries was summed over twenty years assuming a consistent birth rate during this period and applying a discount rate of 3.5%. The final discounted patient population who could potentially benefit from postnatal depression screening was estimated to equal 10,080,946.

Table 6.5: Illustration of EVPI Calculation

Draw Number	Net Benefit A	Net Benefit B	Net Benefit C	Best Strategy	Maximum Net Benefit*
1	10	9	8	A	10
2	11	11	9	A and B	11
3	8	7	11	C	11
4	9	13	8	B	13
5	12	10	7	A	12
6	10	14	6	B	14
7	8	13	12	B	13
8	9	12	8	B	12
9	9	7	9	A and C	9
10	10	13	8	B	13
Mean Net Benefit	9.6	**10.9****	8.6		**11.8*****
EVPI*****					**11.8-10.9**=0.9

Notes: The table demonstrates how an EVPI is calculated for a *hypothetical* PSA conducted using 10 random draws for strategies A, B and C.
* The maximum net benefit per draw is calculated as the net benefit associated with the best strategy.
**The largest mean net benefit given current information is calculated by estimating the mean net benefit for each strategy across *all* draws and identifying the largest of these values.
***The mean maximum net benefit given perfect information is calculated by summing the maximum net benefits across all draws and dividing by the number of draws (i.e. 118/10).
****The EVPI (per individual) is calculated by subtracting B from C.

6.3.5.4 Value of Information Analysis: EVPPI

As well as the overall EVPI, value of information analysis can identify the expected value of partial perfect information (EVPPI). Specifically, the EVPPI identifies the maximum monetary benefit that would be achieved by reducing the uncertainty associated with individual parameters or small groups of parameters. Therefore, the EVPPI places an upper limit on the cost of future research specific to these individual parameters or parameter groups (Ades et al., 2004). Calculation of the EVPPI can inform future research objectives by indicating the type of additional evidence likely to be most useful for future decisions, and equally the type of evidence likely to be least useful (Drummond et al., 2015).

The calculation of EVPPI followed guidance by Ades et al. (2004), Brennan et al. (2007) and Heath et al. (2016), requiring the PSA to be re-conducted in the following stages:

(1) The parameter of interest Φ was identified and PSA was conducted over one draw to identify a random variable for Φ equal to φ. That is, the first stage in the EVPPI calculation identified $\Phi = \varphi$.
(2) PSA was conducted as previously described in Section 6.3.5.2 across n random draws. This excluded the parameter of interest Φ, whose value was held equal the value of the single random variable φ. That is, $\Phi = \varphi$ for all n draws whilst the value of other model parameters varied across individual probabilistic draws.

(3) The largest mean net benefit given current information (NB_{ci}) and the mean maximum net benefit given perfect information (NB_{pi}) were calculated as described in Section 6.3.5.3 and Table 6.5

(4) Procedures (1) – (3) were repeated over N iterations with each iteration using a different fixed value of φ for Φ in the PSA (i.e. $\Phi = \varphi_1\ \varphi_2, .., \varphi_N$). Each iteration identified N distinct NB_{ci} and NB_{pi} values.

(5) The mean NB_{ci} and mean NB_{pi} was obtained across all iterations by summing the distinct NB_{ci} and NB_{pi} values for $\varphi_1, \varphi_2, \ldots, \varphi_N$, and dividing these totals by N.

(6) The final EVPPI was calculated as: mean NB_{pi} – mean NB_{ci}.

The procedure was applied to identify multiple EVPPIs by changing the parameter of interest Φ. The EVPPI was calculated for the following parameter groups: EPDS probabilities; Whooley probabilities; prevalence and risk rates; maternal healthcare costs/utilities; children's combined healthcare costs and QALYs and children's returns to education. The EVPPIs were calculated for cost-effectiveness thresholds where k=£20,000 and k=£30,000 and for v/k=1 when assuming a cross-sectoral decision maker's perspective. Because of large computational requirements, this analysis reduced the PSA to n=1000 simulations, while N=100 iterations were used for Φ.

6.4 Results

6.4.1 Health Centric Perspective: Maternal Outcomes

Strategy descriptions are abbreviated throughout the results. The number in parenthesis () following the strategy description indicates the cut-off point used for that strategy for instance EPDS (16) is the EPDS at a cut-off of 16. The first section of the results addresses research question one (6.2.3) assessing the cost-effectiveness of screening for postnatal depression *without* including effects to children, this being the approach adopted by NICE (2018).

The cost-effective strategy for cost-effectiveness thresholds k=£20,000 and k=£30,000 was the administration of the Whooley + EPDS (16). This strategy dominated standard care case (standard GP care) identification, having lower mean costs and higher mean QALYs. The probability the strategy was cost-effective equalled 0.53 at k=£20,000. This probability decreased upon increasing the threshold value and was equal to 0.24 when k=£30,000.

Table 6.6: Mean Cost-Effectiveness Results for Health Centric Decision Perspective

Strategy	Mean QALYs	Mean Healthcare Costs	ICER[1]	Probability Cost-effective (k=£ 20,000)	Probability Cost-effective (k=£30,000)
Maternal Outcomes Only[2]					
Whooley + EPDS (12)	0.84897	£92.00	£37,621	0.102	0.167
Whooley + EPDS (14)	0.84875	£83.78	£32,137	0.076	0.087
Whooley + EPDS (16)	0.84838	£71.94		0.528	0.236
Maternal & Child Outcomes[3]					
Whooley + EPDS (11)	0.8483	£100.51	£520,564	0.096	0.137
Whooley + EPDS (12)	0.8483	£97.58	£30,914	0.107	0.141
Whooley + EPDS (14)	0.8480	£89.67	£26,818	0.088	0.084
Whooley + EPDS (16)	0.8476	£78.28		0.427	0.188

Notes: Strategies are ranked by mean healthcare costs (highest to lowest). Results are only reported for non-dominated/extendedly dominated strategies.
1: ICERs were calculated vs the next non-dominated higher costing strategy.
2: Cost-effectiveness results for maternal health outcomes only.
3: Cost-effectiveness results for maternal and child health outcomes.

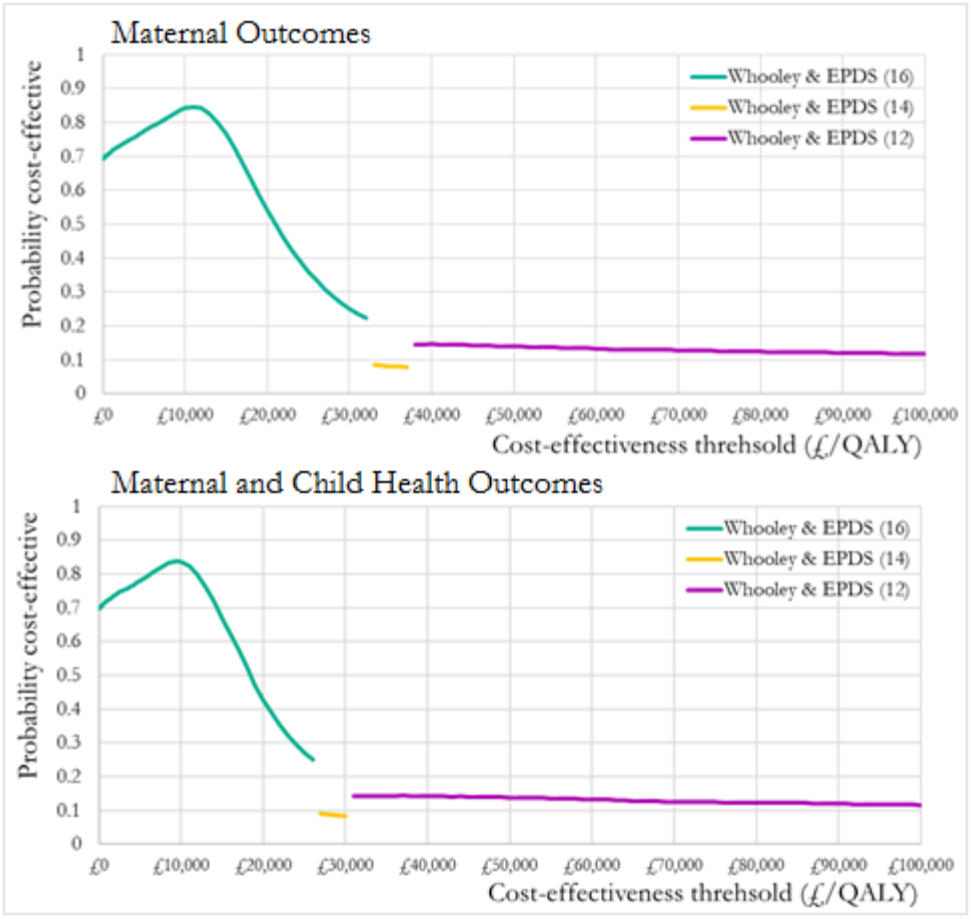


Figure 6.8 Cost-effectiveness acceptability frontiers for maternal vs. maternal & child health outcomes

The cost-effective strategy was dependent on the lower bounds of the cost-effectiveness threshold. As indicated by the ICERs in Table 6.6: the Whooley + EPDS (14) was cost-effective at k=£32,137; while the Whooley +EPDS (12) was cost-effective with k>£37,621. The cost-effective strategy and its associated probability of cost-effectiveness is displayed for thresholds from £0-£100,000 in the cost-effectiveness acceptability frontier in Figure 6.8. Decisions were most uncertain at higher cost-effectiveness thresholds where the probability of cost-effectiveness for the Whooley + EPDS (12) plateaued at roughly p=0.12.

All strategies involving the PHQ-9, case identification through standard care, and sole administration of the EPDS were dominated at all cost-effectiveness thresholds and had a probability of being cost-effective close or equal to zero. Table 6.6 reports results for all non-dominated and extendedly dominated strategies. Full results across all strategies are available in Appendix 6.1.

6.4.2 Health Centric Perspective: Maternal and Child Outcomes

The second question expanded the analytical perspective beyond that addressed in the NICE (2018) analysis, establishing the cost-effective strategy when including the lifetime QALYs and healthcare costs for children exposed to postnatal depression symptoms. The inclusion of child health outcomes resulted in a change in the cost-effective strategy when k=£30,000, where a more sensitive and less specific strategy, the Whooley + EPDS (14), became the cost-effective option (p=0.08). However, the change in the cost-effective strategy was dependent on the value of the cost-effectiveness threshold as the Whooley + EPDS (16) remained cost-effective for k=£20,000 (p=0.43), Table 6.6. All strategies previously dominated or extendedly dominated for the maternal only perspective remained so. This excluded the Whooley + EPDS (11) strategy, which had an ICER of £520,564 far beyond the recommended cost-effectiveness threshold (NICE, 2013).

The PSA produced similar results for both the maternal only and maternal & child health centric decision perspectives: As illustrated in Figure 6.8, the inclusion of child QALYs and healthcare costs appeared to shift the CEAF to the left. This resulted in increased decision uncertainty between cost-effectiveness thresholds of k=£20,000 and k=£30,000. Full cost-effectiveness and PSA results for the maternal and child health centric decision perspective are described in Appendix 6.2.

6.4.3 Cross-Sectoral Perspective: Maternal and Child Outcomes

The third research question established cost-effectiveness for a cross-sectoral decision perspective by including lifetime returns in the education sector as decision endpoint alongside maternal and child QALYs & healthcare costs. Firstly, a cross-sectoral analysis was conducted across all not dominated or extendedly dominated strategies from the previous section (6.4.2), indicating whether returns in the education sector were positive or negative when compared with standard care case identification. The EPDS strategies with high specificity (Whooley + EPDS (16) and Whooley + EPDS (14) had a negative incremental net benefit in the education sector, whereas strategies with more balanced sensitivity and specificity (Whooley + EPDS (12) and Whooley + EPDS (11)) had positive incremental returns in the education sector, Table 6.7.

Table 6.7: Mean Cost-Effectiveness Results for Cross-Sectoral Decision Perspective

Strategy	Net Benefit in Education	Incremental Returns in Education	NCB[1] k=£10,000	NCB[1] k=£20,000	NCB[1] k=£30,000
Whooley + EPDS (11)	Positive	£7.68	24.65	54.18	83.70
Whooley + EPDS (12)	Positive	£5.96	**25.80**	**55.27**	**84.74**
Whooley + EPDS (14)	Negative	-£0.15	25.04	51.95	78.86
Whooley + EPDS (16)	Negative	-£9.29	23.04	45.70	68.37

Notes: 1: NCB= Net Consumption Benefit
Results for all non-dominated/ extendedly dominated strategies. All net benefits are calculated vs. standard care case identification. The cost-effective strategy is indicated in **bold**.

Secondly, cost-effectiveness was more formally examined for the cross-sectoral perspective by estimating the incremental net consumption benefit of each strategy versus standard care, assuming consumption in the health and education sector is valued equivalently (v/k=1). The inclusion of children's returns in education resulted in a change in the cost-effective strategy where a more sensitive and less specific strategy became cost-effective: the Whooley + EDPS (12) strategy was consistently cost-effective, having the highest incremental net health benefit at thresholds of k=£10,000, k=£20,000, and k=£30,000, Table 6.7.

The cost-effective outcome was robust for most consumption ratios in health and education (v/k) where the Whooley + EPDS (12) remained cost-effective. The cost-effective strategy only changed once v/k was: reduced to 0.5 or lower where the Whooley + EPDS (10) strategy became cost-effective for both k=£20,000 and k=£30,000; or increased to 4 when where the Whooley + EPDS (16) strategy became cost-effective for k=£20,000. Figure 6.9A and 6.9B depict incremental net benefits of each strategy vs. standard care across a range of v/k ratios for k=£20,000 and k=£30,000.

	v	£5,000	£10,000	£13,333	£15,000	£20,000	£26,667	£30,000	£40,000	£80,000
	k	£20,000	£20,000	£20,000	£20,000	£20,000	£20,000	£20,000	£20,000	£20,000
	v/k	0.25	0.50	0.67	0.75	1.00	1.33	1.50	2.00	4.00
EPDS (cutoff 7)		-14.25	-46.03	-67.22	-77.81	-109.59	-151.97	-173.15	-236.72	-490.96
EPDS (cutoff 8)		-5.20	-27.93	-43.09	-50.67	-73.41	-103.73	-118.88	-164.36	-346.26
EPDS (cutoff 9)		1.16	-12.64	-21.84	-26.44	-40.24	-58.64	-67.83	-95.43	-205.82
EPDS (cutoff 10)		4.40	-4.60	-10.60	-13.60	-22.60	-34.61	-40.61	-58.61	-130.61
EPDS (cutoff 11)		5.76	2.27	-0.06	-1.23	-4.72	-9.38	-11.71	-18.69	-46.64
EPDS (cutoff 12)		6.17	4.90	4.05	3.62	2.35	0.65	-0.20	-2.75	-12.94
EPDS (cutoff 13)		4.88	3.13	1.97	1.39	-0.36	-2.69	-3.85	-7.35	-21.32
EPDS (cutoff 14)		2.48	3.95	4.93	5.42	6.89	8.85	9.83	12.76	24.52
EPDS (cutoff 15)		-1.70	1.79	4.12	5.28	8.77	13.42	15.75	22.73	50.66
EPDS (cutoff 16)		-3.95	0.70	3.80	5.36	10.01	16.22	19.32	28.63	65.88
PHQ9 (cutoff 9/10)		3.46	-3.55	-8.23	-10.57	-17.58	-26.93	-31.61	-45.64	-101.74
Whooley + EPDS (cutoff 7)		17.60	19.66	21.03	21.72	23.78	26.53	27.90	32.02	48.50
Whooley + EPDS (cutoff 8)		20.78	26.00	29.49	31.23	36.45	43.42	46.91	57.36	99.16
Whooley + EPDS (cutoff 9)		21.35	29.60	35.10	37.84	46.09	57.09	62.59	79.09	145.08
Whooley + EPDS (cutoff 10)		**21.48**	**31.34**	37.92	41.21	51.07	64.23	70.80	90.53	169.45
Whooley + EPDS (cutoff 11)		19.30	30.93	38.68	42.55	54.18	69.67	77.42	100.67	193.66
Whooley + EPDS (cutoff 12)		18.28	30.61	**38.83**	**42.94**	**55.27**	**71.70**	**79.92**	**104.58**	203.19
Whooley + EPDS (cutoff 13)		17.31	29.44	37.53	41.57	53.70	69.87	77.95	102.21	199.24
Whooley + EPDS (cutoff 14)		12.87	25.90	34.58	38.92	51.95	69.31	78.00	104.05	208.25
Whooley + EPDS (cutoff 15)		7.43	20.91	29.90	34.40	47.88	65.85	74.84	101.80	209.64
Whooley + EPDS (cutoff 16)		4.46	18.21	27.37	31.96	45.70	64.04	73.20	100.70	**210.70**
Whooely + PHQ9 (cutoff 9/10)		19.24	29.64	36.58	40.05	50.45	64.32	71.26	92.07	175.30
Standard Care		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Key
Lowest Net Benefit
Highest Net Benefit

Figure 6.9A: Cost-effectiveness results when including children's QALYs, healthcare costs and education costs assuming a cost-effectiveness threshold of k=£20,000. Results are obtained across a range of feasible consumption values for health (v). Incremental net consumption benefits were calculated for all strategies vs. standard care case identification. All child outcomes are discounted at 3.5%.

	v	£7,500	£15,000	£20,000	£22,500	£30,000	£40,000	£45,000	£60,000	£120,000
	k	£30,000	£30,000	£30,000	£30,000	£30,000	£30,000	£30,000	£30,000	£30,000
	v/k	0.25	0.50	0.67	0.75	1.00	1.33	1.50	2.00	4.00
EPDS (cutoff 7)		-16.99	-51.51	-74.52	-86.02	-120.54	-166.57	-189.58	-258.61	-534.76
EPDS (cutoff 8)		-4.96	-27.47	-42.47	-49.97	-72.48	-102.49	-117.49	-162.50	-342.55
EPDS (cutoff 9)		3.96	-7.04	-14.38	-18.04	-29.04	-43.71	-51.04	-73.05	-161.05
EPDS (cutoff 10)		8.55	3.70	0.47	-1.15	-6.00	-12.47	-15.70	-25.40	-64.19
EPDS (cutoff 11)		11.13	13.01	14.26	14.88	16.76	19.26	20.51	24.26	39.27
EPDS (cutoff 12)		12.01	16.58	19.62	21.14	25.71	31.80	34.84	43.98	80.51
EPDS (cutoff 13)		10.45	14.27	16.82	18.10	21.92	27.02	29.56	37.21	67.79
EPDS (cutoff 14)		8.30	15.60	20.46	22.90	30.19	39.92	44.78	59.37	117.74
EPDS (cutoff 15)		3.91	13.00	19.07	22.10	31.20	43.33	49.40	67.59	140.38
EPDS (cutoff 16)		1.55	11.71	18.48	21.87	32.03	45.58	52.35	72.67	153.95
PHQ9 (cutoff 9/10)		7.87	5.26	3.52	2.65	0.05	-3.43	-5.17	-10.38	-31.24
Whooley + EPDS (cutoff 7)		22.96	30.38	35.33	37.80	45.22	55.12	60.06	74.90	134.27
Whooley + EPDS (cutoff 8)		27.18	38.80	46.56	50.43	62.06	77.56	85.31	108.56	201.58
Whooley + EPDS (cutoff 9)		28.39	43.68	53.88	58.98	74.27	94.66	104.86	135.44	257.79
Whooley + EPDS (cutoff 10)		**28.84**	**46.07**	57.56	63.30	80.53	103.50	114.99	149.44	287.27
Whooley + EPDS (cutoff 11)		26.69	45.69	58.36	64.70	83.70	109.04	121.71	159.72	311.77
Whooley + EPDS (cutoff 12)		25.65	45.35	**58.48**	**65.04**	**84.74**	**111.00**	**124.13**	**163.51**	**321.07**
Whooley + EPDS (cutoff 13)		24.50	43.82	56.70	63.14	82.46	108.23	121.11	159.75	314.31
Whooley + EPDS (cutoff 14)		19.60	39.35	52.52	59.10	78.86	105.19	118.36	157.87	315.89
Whooley + EPDS (cutoff 15)		13.47	32.99	46.01	52.51	72.03	98.06	111.07	150.11	306.27
Whooley + EPDS (cutoff 16)		10.12	29.54	42.48	48.95	68.37	94.26	107.20	146.03	301.35
Whooely + PHQ9 (cutoff 9/10)		26.39	43.94	55.64	61.49	79.04	102.44	114.13	149.23	289.63
Standard Care		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Key
- Lowest Net Benefit
- Highest Net Benefit

Figure 6.9B: Cost-effectiveness results when including children's QALYs, healthcare costs and education costs assuming a cost-effectiveness threshold of k=£30,000. Results are obtained across a range of feasible consumption values for health (v). Incremental net consumption benefits were calculated for all strategies vs. standard care case identification. All child outcomes are discounted at 3.5%.

The inclusion of costs incurred in the education sector appeared to decrease the overall decision uncertainty when compared with the analyses excluding these costs, as illustrated in Figure 6.10 which presents the cost-effectiveness acceptability frontier for consumption ratios v/k=1. The probability of the cost-effective strategy being the correct decision did not exceed 0.5 and was less than 0.1 when cost-effectiveness thresholds ranged from £10,000-£30,000.

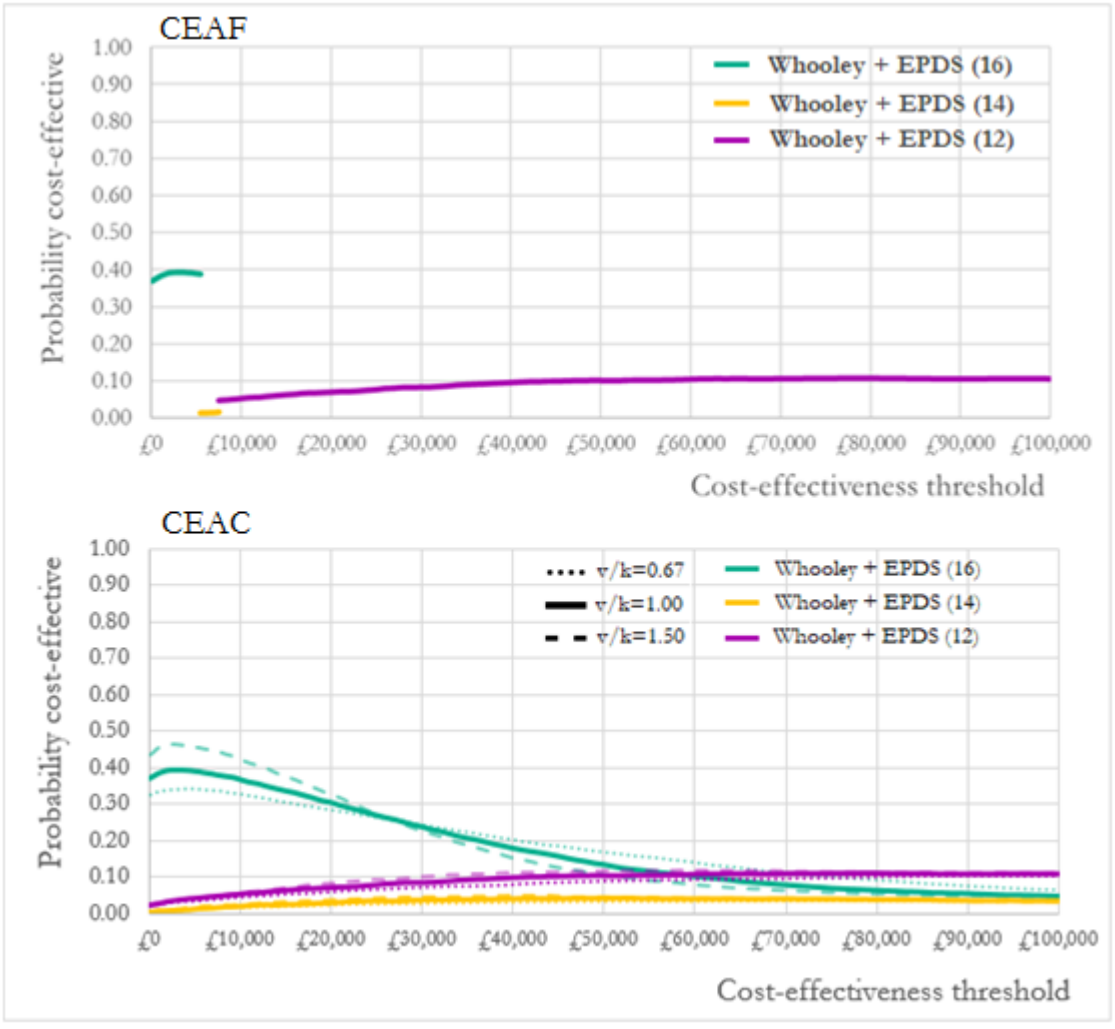


Figure 6.10: Probabilistic sensitivity analysis results for the cross-sectoral perspective which includes children's lifetime QALYs, healthcare costs and returns to education alongside maternal decision endpoints (all discounted at 3.5%)
Above: A cost-effectiveness acceptability frontier where children's returns to education costs are included in the decision problem using a ratio of v/k=1.
Below: Cost effectiveness acceptability curves demonstrating the variability in probabilities of cost-effectiveness for non-dominated/extendedly dominated strategies according to different v/k ratios.

Different levels of uncertainty were identified for different v/k ratios at the same cost-effectiveness thresholds. For instance, at k=£20,000 the Whooley + EPDS 16 strategy had a probability of being cost effective of p=0.3 when v/k=1.5, p=0.28 when v/k=1 and p=0.32 when v/k=0.67. As illustrated by the cost-effectiveness acceptability curves in Figure 6.10, each strategy's probability of cost-effectiveness followed similar trajectories for each v/k ratio and differences were generally minimal when ratios which ranged from 0.67-1.50. Differences were most substantial for lower cost-effectiveness thresholds where k<£10,000.

6.4.4 Deterministic Sensitivity Analyses

The results of the deterministic sensitivity analysis are summarised in Table 6.8 and reported in full in Appendix 6.4. The scenario with the largest impact on the cost-effectiveness decision occurred when assuming false positive mothers incurred no reduction in utility rather than the 2% reduction assumed in the base case. Under this assumption several previously dominated became non-dominated as a result of increases in their associated mean QALYs. This led to standard care case identification being identified as cost-effective at thresholds of k<£27,944, whilst the Whooley + EPDS (11) strategy was cost-effective when k=£30,000.

The application of a 1.5% rather than a 3.5% discount rate to children's lifetime QALYs *and* costs reduced the ICERs between all non-dominated or extendedly dominated strategies. The change in discount rate affected the cost-effective strategy. At a threshold of k=£30,000: the Whooley + EPDS (12) was cost-effective when accounting for children's lifetime QALYs and healthcare costs (Table 6.8); meanwhile the Whooley + EPDS (10) became cost-effective for the cross-sectoral decision perspective (Appendix 6.4F). The scenario which discounted QALYs (1.5%) and costs (3.5%) at different rates influenced the cost-effectiveness result for the health-centric decision perspective where the Whooley + EPDS 12 strategy became cost-effective but had no effect for the cross-sectoral decision perspective (Appendix 6.4G).

Similarly, changes to the prevalence rate of postnatal depression resulted in reduced ICERs between the non-dominated or extendedly dominated strategies. For the perspective which included maternal health outcomes, and when assuming a postnatal depression population prevalence of 20% rather than 8.7%, the cost-effective strategy changed to the Whooley + EPDS (12) for k=£30,000. This finding was dependent on the value of the cost-effectiveness threshold as the Whooley + EPDS 16 remained cost-effective for a 20% prevalence rate when k=£20,000.

Table 6.8: Deterministic Sensitivity Analysis Mean Cost-Effectiveness Results

Strategy	Mean QALYs	Mean Costs	ICERs*
Scenario 1[B]: Discount Rate= 1.5%			
Whooley + EPDS (10)	0.8474	£116.05	£251,856
Whooley + EPDS (11)	0.8474	£109.47	£34,386
Whooley + EPDS (12)	0.8473	£106.69	£24,109
Whooley + EPDS (14)	0.8470	£99.27	£21,213
Whooley + EPDS (16)	0.8465	£88.63	
Scenario 2[A]: Postnatal Depression Prevalence=20%			
Whooley + EPDS (9)	0.8356	£195.42	£998,450
Whooley + EPDS (10)	0.8356	£189.62	£40,260
Whooley + EPDS (11)	0.8353	£177.28	£37,522
Whooley + EPDS (12)	0.8352	£171.96	£23,544
Whooley + EPDS (14)	0.8345	£155.25	£22,403
Whooley + EPDS (16)	0.8334	£130.64	
Scenario 3[A]: Utility reduction for false positive Mothers = 0%			
EPDS (8)	0.8500	£175.76	£527,538
Whooley + EPDS (8)	0.8499	£113.15	£48,303
Whooley + EPDS (9)	0.8498	£106.08	£45,033
Whooley + EPDS (10)	0.8497	£102.10	£30,213
Whooley + EPDS (11)	0.8495	£95.01	£29,322
Whooley + EPDS (12)	0.8494	£92.00	£27,945
Standard Care	0.8490	£82.07	£18,288
Whooley + EPDS (16)	0.8484	£71.94	
Scenario 4[B]: Children's Exposure Proportionate to Depression Duration			
Whooley + EPDS (11)	0.8482	£101.43	£6,297,315
Whooley + EPDS (12)	0.8482	£98.46	£33,918
Whooley + EPDS (14)	0.8479	£90.40	£29,712
Whooley + EPDS (16)	0.8475	£78.79	

Notes: Results reported for all non-dominated or extendedly dominated strategies. All outcomes are discounted at 3.5% unless otherwise stated.
(A) Health centric decision perspective with maternal outcomes only.
(B) Health centric decision perspective with maternal and child outcomes.
*ICERs calculated vs the next non-dominated, higher costing strategy.

Lifetime effects were estimated for the base case by assuming that the incremental effects of postnatal depression on children's development remained constant beyond the age of eleven (the final age at which effects were observed in chapter three). This assumption was investigated for the health centric decision perspective in two scenario analyses by applying a 25% decrease and 25%

increase to the incremental effects to approximate diminishing and snowballing effect sizes over time. The scenario which applied a 25% diminishing effect sizes increased the value of the ICERs for all non-dominated or extendedly dominated strategies but did not influence the cost-effectiveness result, where the Whooley + EPDS (16) strategy and Whooley + EPDS (14) strategy remained cost-effective when k=£20,000 and k=£30,000. In contrast, the scenario applying a 25% snowballing effect size decreased the value of the ICERs which changed the cost-effectiveness result, where the Whooley + EPDS (12) strategy became cost-effective for k=£30,000, Appendix 6.4D.

Finally, the assumption that children were not exposed to postnatal depression symptoms if their mothers responded to treatment was relaxed by assuming that these children incurred lifetime effects proportionate to the time their mothers were in the depressed disease state. This scenario had little impact on the decision and produced the same cost-effective strategies as the base case analysis at thresholds of k=£20,000 (Whooley + EPDS (16)) and k=£30,000 (Whooley + EPDS (14)).

6.4.5 Value of Information Analysis

6.4.5.1 Expected Value of Perfect Information

The base case value of information analysis was conducted for the model including child healthcare costs and QALYs. At a cost-effectiveness threshold of k=£20,000, the base case EVPI was £3.50 per patient and £35,326,497 per population, which increased to £16.20 per patient and £163,264,542 per population when the threshold was raised to k=£30,000. There was a continued (linear) increase in the EVPI for thresholds above k=£30,000 which occurred as a result of a consistently high level of uncertainty associated with the Whooley + EPDS (12). In contrast, at lower cost-effectiveness thresholds (k<£20,000) the EVPI was close to zero due to the high probability of the Whooley + EPDS (16) being the cost-effective strategy, Figure 6.11.

The inclusion of children's educational outcomes resulted in substantially larger EVPIs which were equal to: £29.05 (person) and £292,901,679 (population) for k=£20,000 and v/k=1; and £32.83 (person) £330,938,234 (population) for k=£30,000 and v/k=1. As illustrated in Figure 6.11, the EVPI increased (decreased) for all cost-effectiveness thresholds when v/k=0.67 (v/k=1.5) where it was assumed that consumption in the education sector is of a higher (lower) value than consumption in the health sector.

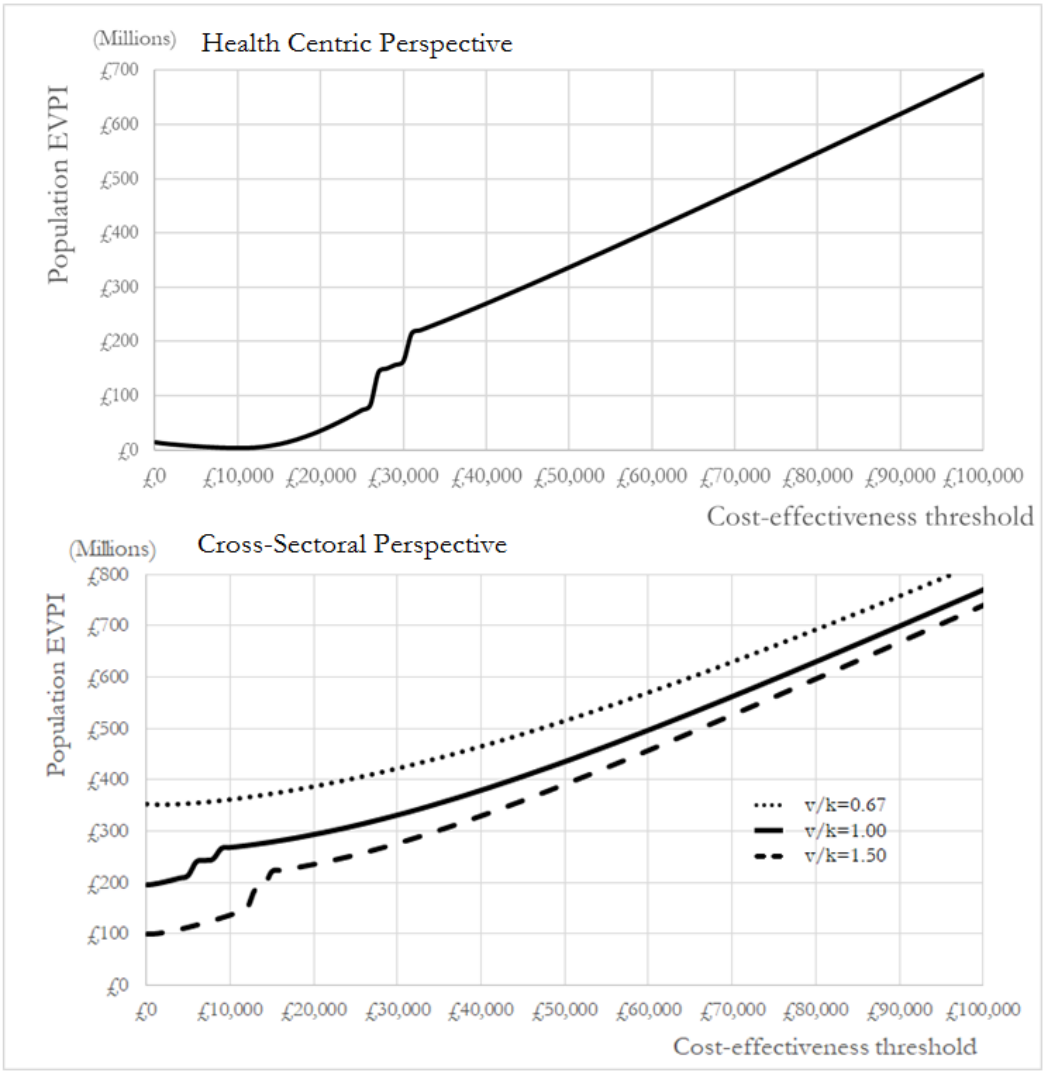


Figure 6.11: The population EVPI at different cost-effectiveness threshold for models with both maternal and child outcomes discounted at 3.5% and for an assumed patient population in England and Wales= 10,080,946: Above results are obtained for a health centric perspective (children's QALYs and healthcare costs). Below results are obtained for a cross sectoral perspective (children's QALYs, healthcare costs and returns to education) applying different v/k ratios.

6.4.5.2 Expected Value of Partial Perfect Information

The population EVPPIs for all grouped parameters are illustrated in Figure 6.12 for cost-effectiveness thresholds ranging from £0 to £50,000. Large population EVPPIs were associated with the group of EPDS model parameters (sensitivities and specificities) which were equal to £23,394,201 at a cost-effectiveness threshold of k=£20,000 and increased to £139,084,128 when

k=£30,000. This contrasted with the group of parameters estimating child healthcare costs and QALYs which had lower EVPPIs equal to zero for k=£20,000 and £44,570,493 when k=£30,000. Similarly, EVPPIs increased as cost-effectiveness thresholds increased for maternal QALYs (£26,187 (k=£20,000), £66,112,691 (k=£30,000)) and for the Whooley questions sensitivity & specificity parameters (£0.00 (k=20,000); £7,578,073 (k=£30,000)).

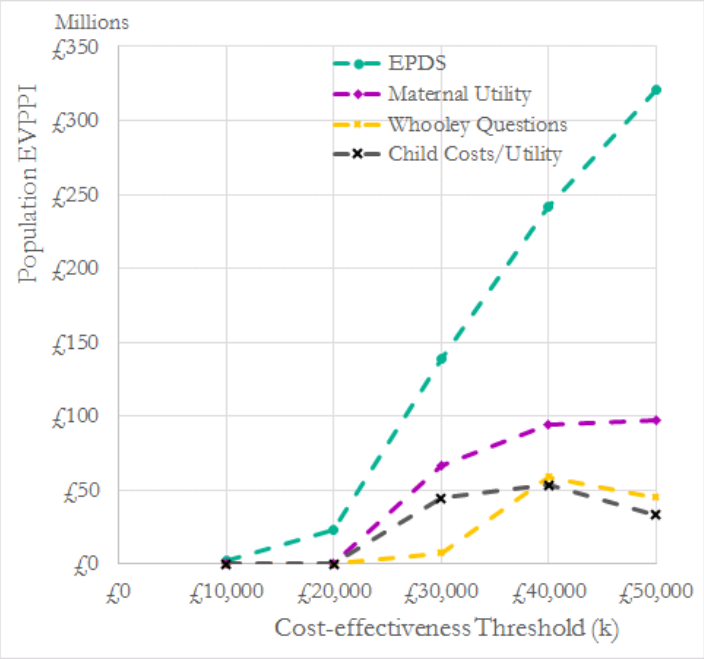


Figure 6.12: Population EVPPI across cost-effectiveness thresholds £0-£50,000 for parameter groups (a) EPDS sensitivities and specificities; (b) maternal utility; (c) Whooley questions sensitivity and specificity; and (d) child lifetime QALYs and healthcare costs.

6.5 Discussion

6.5.1 Cost-Effectiveness Results

6.5.1.1 General Cost-Effectiveness Results

This analysis identified a dual screen using the Whooley questions followed by a formal screening strategy (EPDS/PHQ-9) for all positive initial screens as the general cost-effective case identification strategy for postnatal depression in the UK. The strategy consistently outperformed

all the other general strategies (standard care case identification, and single screening strategies with the EPDS/PHQ-9) from three different decision-making perspectives, across a range of cost-effectiveness thresholds and for a variety of different scenario analyses. These results are consistent with current NICE (2018) recommendations which suggest that GPs should consider posing the Whooley question to mothers during their first contact with primary care after child birth and should follow-up with either an EPDS or PHQ-9 screen if women respond positively to the initial screen.

6.5.1.2 Cost-Effectiveness: Maternal Outcomes Only

When adopting a perspective limited to maternal QALYs/healthcare costs, this analysis identified the cost-effective screening strategy as the Whooley questions followed by the EPDS with a threshold equal to 16. Whilst the results are consistent with NICE (2018) recommendations, the specific cost-effectiveness results diverge from those obtained in the NICE (2018) economic evaluation. NICE (2018) identified the PHQ-9 as the cost-effective second screening instrument, whilst this research identified the EPDS as cost-effective. The difference in results is likely to have occurred as this analysis investigated the EPDS across a range of threshold values, in contrast to the single threshold investigated by NICE (2018).

Whilst NICE (2018) identified a different instrument as cost-effective for the second screen, their results are consistent with this analysis in the sense that the cost-effective screen is that with the highest specificity. The cost-effective strategy in this analysis was the EPDS with a threshold of 16 which had the highest associated specificity (0.99) and lowest sensitivity (0.28); meanwhile NICE (2018) identified the PHQ-9 with a threshold of 9 (specificity=0.88, sensitivity=0.75) as cost-effective when compared to the EPDS with a cut-off of 9/10 (specificity=0.85, sensitivity=0.83). As this analysis utilised the same modelling structure and assumption as NICE (2018), it is likely the highly specific EPDS threshold would have been cost-effective in the NICE (2018) model had this strategy been investigated. These conclusions are supported by Paulden et al. (2009) who found the most specific threshold to be the cost-effective EPDS strategy when screening for postnatal depression, although all single screening strategies were dominated by standard care case identification.

The impact of employing a highly sensitive screen (the Whooley questions) followed by a highly specific EPDS screen would result in treatments being targeted towards postnatal women with a very high probability of being depressed, thus minimising the costs and disutility occurring through incorrect treatment assigned to non-depressed women. However, the adoption of a high EPDS

threshold is also associated with a reduction in test sensitivity and therefore a reduction in the proportion of truly depressed women receiving treatment. This would be expected to equal approximately 33% for the EPDS 16 threshold, which is substantially less than the 54% of truly depressed women who would receive treatment under standard care case identification. Whilst not providing treatment to depressed women might seem counterintuitive and/or unethical, the negative effects are arguably outweighed by the benefits of not providing costly (and potentially harmful) treatments for women who do not need them – these resources might be better allocated elsewhere in the healthcare system.

The results of the deterministic sensitivity analysis further illustrate how cost-effectiveness results were driven by the balance between providing beneficial treatment for true positive mothers and not providing costly/harmful treatment to false positive mothers. Where wrongly assigned treatment was assumed to only result in additional healthcare costs (and not a 2% disutility), the cost-effectiveness result moved towards less specific and more sensitive strategies (i.e. standard care case identification for k<£27,944 and the Whooley + EPDS (11) for k=£30,000). The change in cost-effectiveness occurred as more sensitive screening strategies are assumed to provide treatment for a higher proportion of depressed *and* non-depressed women than more specific strategies. The removal of the 2% disutility would mean that it is less important to avoid treating non-depressed women, thus allowing for screening strategies with relatively higher levels of overall treatment provision.

6.5.1.3 Cost-Effectiveness: Maternal and Child Outcomes

The inclusion of children's lifetime decision endpoints had an influence on the cost-effectiveness results. When children's QALYs and healthcare costs were included in the decision-making perspective at a willingness-to-pay of £30,000 per QALY, the cost-effective strategy changed to the less specific and more sensitive EPDS with a cut-off equal to 14. This recommendation might be more comfortable for policy makers as it would result in a similar proportion of truly depressed women being allocated treatment (53%) as in standard care case identification. The change in decision for this perspective is driven by the additional treatment benefits in children that were assigned to mothers who recovered from depression – these outweighing some of the costs/disutility associated with treating false positive cases.

No change in the cost-effective strategy occurred for cost-effectiveness thresholds less than or equal to £20,000 per QALY: decreasing the value placed on each QALY resulted in less weighting

being applied to children's lifetime health benefits which no longer outweighed the costs/disutility of treating false positive mothers.

The influence of child outcomes on the cost-effective strategy was more pronounced when the decision perspective was expanded to also include children's lifetime returns to education. The cost-effective strategy changed at both cost-effectiveness thresholds (k=£20,000, k=£30,0000), where the Whooley questions followed by the EPDS with a threshold of 12 became the recommended screening strategy. This was a more sensitive (0.68) and less specific (0.93) cost-effective strategy than for the previous two perspectives and would be expected to result in the treatment of a greater proportion (73%) of truly depressed women than would be treated under standard care (54%). The change in policy recommendation occurs for the reason discussed above. The addition of educational outcomes attached further additional benefits to successful treatment of postnatal depression, making it more acceptable to incur costs through treatments wrongly assigned to non-depressed women.

Decision makers might have more confidence in the cross-sectoral results in this analysis given the cost-effective strategy remained robust across all v/k ratios. Whilst the consumption values for health and education are not known, it is unlikely that consumption (£s spent) in the health sector is valued four times higher or lower than consumption in the education sector. In addition, the results of this analysis can be interpreted to predict cost-effectiveness *if* it were assumed that the v/k ratios extended beyond the range investigated. Intuitively if less value is placed on consumption in education (v/k<1), the decision more closely resembles the health centric decision perspective thus tending towards a cost-effective with the EPDS cut-off equal to 14 at the lower v/k limit. In contrast, when greater value was placed on consumption in education (v/k>1), the cost-effective strategy tends towards strategies with higher sensitivities with an upper limit likely to be the EPDS strategy with the highest non-dominated ICER (threshold equal to 10).

The results from both health centric and cross-sectoral decision perspectives also remained robust when applying different assumptions to child related model parameters. Informed by results from chapter three, the original model assumed the incremental lifetime effects of postnatal depression on child development continued as a consistent effect across the lifespan. Children's lifetime outcomes still influenced cost-effective decision when assuming 25% of this effect was lost as children aged. Similarly, the influence of child lifetime parameters on the cost-effectiveness result were maintained when assuming mothers who were successfully treated for postnatal depression transmitted some effects to children during the time they were depressed.

In contrast, results were affected when changing the magnitude of the discount rate applied to children's lifetime outcomes from 3.5% to 1.5%. The lower discount rate resulting in less specific and more sensitive EPDS strategies with lower thresholds becoming cost-effective. Changes to cost-effectiveness occurred as lower discount rates are associated with a larger magnitude of benefit for the lifetime outcomes, therefore applying more weighting to the benefits of treating true positive mothers. The effect of discounting was more pronounced for the perspective including children's lifetime returns to education. These outcomes were larger in magnitude and had more influence on the overall cost-effectiveness when compared with children's lifetime health decision endpoints.

Whether decision makers should include costs and benefits related to children remains a controversial question. Drummond et al. (2015) suggests economic evaluation ought to identify *all* the *important* consequences associated with health technologies. Therefore, it is often sufficient to establish costs and benefits occurring only for the direct recipient of the health technology, assuming the effects on non-recipients are inconsequential. Ride (2018) indicates that non-patient effects have begun to be considered by key decision makers, like NICE, in the UK, stating several examples including economic evaluations of intervention for dementia which account for health effects in carers, and child vaccination programmes which have far reaching benefits in terms of reducing the spread of disease.

There is, therefore, a strong argument for including children's outcomes in policy decisions related to postnatal depression and its treatment. Empirical and theoretical evidence suggests a causal link between maternal depression and child development,[11] and this analysis further demonstrates the influence and importance of child outcomes on these policy decisions.

6.5.1.4 Cost-Effectiveness in Sub-Populations

The deterministic sensitivity analysis suggested it might be appropriate to implement different EPDS strategies for different sub-populations. Lower EPDS thresholds are recommended for populations with a higher prevalence of postnatal depression. The rationale for increasing the EPDS test sensitivity is due to high prevalence populations having a greater proportion of truly depressed mothers, and therefore more profound overall treatment benefits for true positive cases. Separate policies might be advisable for populations with a higher risk of postnatal depression

[11] See evidence reviewed in chapter three.

symptoms than the general population which include low income, poorly educated, and first-time mothers (Goyal et al., 2010).

6.5.2 Probability of Cost-effectiveness

The results of the probabilistic sensitivity analysis indicated a high level of decision uncertainty across all decision perspectives. For instance, when cost-effectiveness thresholds were between £20,000 and £30,000 the largest probability of cost-effectiveness were associated with the Whooley + EPDS 16 strategy which reached a maximum of 0.48 when assessing maternal outcomes only, whilst the cost-effective strategies for the other decision perspectives (Whooley + EPDS 14/12) typically had a probability of cost-effectiveness close to 0.1 throughout this range. The relatively low probability of cost-effectiveness identified for the cost-effective strategies might alarm decision makers. A review by Adalsteinsson and Toumi (2013) suggest that health technologies with associated probabilities of cost-effectiveness less than 40% are rarely recommended by NICE.

However, this economic evaluation assessed cost-effectiveness across a range of similar strategies which only differed in terms of the threshold applied to the EPDS. Consecutive EPDS strategies were only marginally different in terms of their associated specificity/sensitivity. For example, the EPDS with a threshold of 10 had specificity equal to 0.86 and a sensitivity of 0.82, whilst the EPDS with a threshold of 11 had specificity of 0.84 and sensitivity equal to 0.85. Because of the random nature of PSA, marginally different EPDS strategies would be expected to frequently outperform one another during the individual probabilistic draws, thus reducing the probability of cost-effectiveness associated with *individual* EPDS strategies.

If the strategies in this analysis are grouped in terms of a general strategy, then the probability of cost-effectiveness becomes much higher. The general strategy of applying the Whooley questions followed by *any* EPDS threshold in a dual screen was above 95% for all decision perspectives when k=£20,000 or £30,000. If decision makers choose not to adopt one of the dual screening strategies, they would necessarily be choosing to adopt another i.e. standard care case identification. Based on this research the adoption of standard care case identification would almost certainly be the wrong decision as it had an extremely low probability of being cost-effective, equalling close to 1% (k=£20,000) and 0.5% (k=£30,000) across all perspectives.

6.5.3 The Cost-Effectiveness of Conducting Future Research

The relatively high decision uncertainty together with the size of the population potentially benefiting from screening explains the magnitude of the Expected Value of Perfect Information (EVPI). For decision perspectives including maternal and child health endpoints, this reached close to £200 million between cost-effectiveness thresholds of £20,000 and £30,000. The magnitude of the EVPI increased further still when adopting a cross-sectoral perspective. The value of the EVPI is an upper limit on the benefits that might be achieved through future research (Claxton et al., 2002). As even the most expensive trial designs would be unlikely to exceed the identified EVPIs, the value of information analysis indicates future research *has the potential* to be cost-effective.

According to the Expected Value of Partial Perfect Information (EVPPI) it might be most desirable to conduct future research to obtain a more precise estimation of the sensitivities and specificities associated with the EPDS. If this were achieved, then it may be possible to confidently recommend the EPDS at individual thresholds. Given the very large EVPPI, it might be cost-effective to obtain this information through a large randomised trial design despite their high associated costs.

The potential cost-effectiveness of future research regarding the estimation of children's lifetime effects appears to be dependent on the cost-effectiveness threshold. This analysis suggests future research estimating children's lifetime QALYs/healthcare costs is potentially cost-effective for thresholds where k>=£30,000, where the EVPPI was close to £50 million. Again, it is very unlikely that any research costs would exceed these upper bounds, particularly given much of the evidence regarding lifetime effects utilises existing data from longitudinal birth cohorts. However, the decision to conduct future research should also consider the difficulty of obtaining accurate estimates over a long-term time horizon, and consequently the level of precision that might be achieved– i.e. it is likely that future research would only be able to address a small proportion of parameter uncertainty.

The value of information analysis suggests that future research identifying children's lifetime QALYs/healthcare costs does not have the potential to be cost-effective if k=£20,000 or less, where the associated EVPPI was equal to zero. The value of the EVPPI demonstrates the limited influence of children's lifetime healthcare costs and QALYs at these threshold ranges, where the cost-effective strategy remained the same as that identified when assessing maternal health decision endpoints only. The limited influence of children's health outcomes at these thresholds is likely to be due to the relatively small magnitude of effect size and associated confidence intervals for QALY

and healthcare costs endpoints. A full discussion regarding the magnitude of lifetime estimates is provided in chapter seven.

6.5.4 Limitations and Areas of Investigation for Future Research

The priority of this research was to address the overarching aim of the thesis by including estimates of children's lifetime effects in an economic evaluation and discussing their influence on cost-effectiveness results. The considerable time resources required to achieve this objective meant overlooking several limitations that were identified in the decision model.

Firstly, the decision model is likely to have been limited by the time horizon, where maternal depression was only modelled over a 52- week period. There is empirical evidence suggesting that postnatal depression is a risk factor in the development of psychiatric disorders including major depressive disorder (Vliegen et al., 2014). In addition, the results from chapter three suggest substantial effects to child development are likely to occur through accumulated symptoms of maternal depression after the postnatal period.

As discussed above, the cost-effective strategy was sensitive to scenarios where greater weighting was applied to treatment benefits for true positive cases, where more sensitive and less specific EPDS strategies became cost effective. If it is assumed that successful treatments for postnatal depression could reduce the likelihood of future psychiatric problems, then the 52 -week time horizon may underestimate the overall benefits of treatment for postnatal depression. If this were the case, then lower EPDS thresholds (than those recommended here) might become cost-effective. An objective of future research could be the estimation of a decision model assessing the cost-effectiveness of screening for postnatal depression across a longer-term time horizon which could be achieved by applying a Markov model to extend the decision tree (Drummond et al., 2015).

Secondly, apart from the investigation of different EPDS thresholds, this analysis did not attempt to update any of the model parameters utilised in the NICE (2018) model. It was assumed that the parameters were up to date, given NICE (2018) recommendations for postnatal and antenatal care were updated during the same year as this analysis. However, it is possible that more recent data is available for some of the model parameters, for instance, maternal utility for postnatal depression was identified by NICE (2018) in a study by (Sapin et al., 2004), which is unlikely to be the most recent study available for this specific model parameter.

Similarly, this analysis did not attempt to change any NICE (2018) modelling assumptions. For example, the NICE (2018) model assumed a 2% reduction in health utility ought to be applied to false positive women, who received treatment but were not depressed. This assumption was shown to be highly influential on the cost-effectiveness results, where scenarios which applied a disutility equal to zero resulted in a much more sensitive and less specific EPDS strategies becoming cost-effective across all decision perspectives. The influence of this assumption is unsurprising given the drivers of cost-effectiveness observed in this analysis – removing the disutility parameter removes part of the detriment attached to treating false positives.

It would be beneficial for future research to update parameter estimates and address some of the more questionable modelling assumptions applied in the NICE (2018) model. The specific objectives of future research should be informed through the results of both the deterministic and probabilistic sensitivity analyses in this research. It would make most sense for new research priorities to focus on the assumptions shown to be influential on cost-effectiveness, and on model parameters with the largest EVPPIs.

Thirdly, this research was not able to compare cost-effectiveness results for all potential strategies when screening for postnatal depression: There is only evidence regarding the specificity and sensitivity of the PHQ-9 instrument at one threshold value in postnatal women. This strategy instrument was consistently outperformed by at least one EPDS strategy, where 10 different threshold ranges were investigated. As the PHQ-9 is assumed to incur identical administration costs, the results for different EPDS strategies could be applied to the PHQ-9 instrument when additional data becomes available regarding sensitivities/specificities for different cut-off values. That is, policy makers could identify the cost-effectives of the PHQ-9 at a specific threshold by identifying results for the EPDS strategy which corresponded most closely to the estimated sensitivity/specificity).

A fourth limitation may have occurred as a result of the economic evaluation being developed and error checked by a single researcher. Several validity checks were performed at extreme parameter values, confirming that the model results reflected these extreme scenarios. The validity of the model is also enhanced given the similar conclusions that were drawn with existing evidence from NICE (2018) and from Paulden et al. (2009).

Finally, all the limitations regarding the estimation of lifetime child outcomes e.g. causality, imprecision, heteroscedasticity etc. that have been discussed in chapters three and five are applicable to this analysis. The overall validity of estimates obtained from indirect estimation is discussed in detail in chapter seven.

6.5.5 Conclusions

The aims of this chapter are to demonstrate how indirect lifetime effects estimates can be used to inform an applied economic evaluation, and to determine their overall influence on a real -life healthcare allocation decision. The chapter describes the results of an economic evaluation assessing the cost-effectiveness of screening for postnatal depression, for decision perspectives which both included and excluded children's lifetime outcomes.

When considering maternal outcomes only, the results of the analysis suggest it is cost-effective to use a dual case identification strategy of the Whooley questions, followed by the Edinburgh Postnatal Depression Scale (EPDS) at a threshold of 16 for all positive initial screens. In contrast, more sensitive and less specific EPDS cut-off thresholds of 14 or 12 are recommended if the decision maker's perspective is to include child health, or child health and education outcomes. Policy recommendations were influenced as the inclusion of children's lifetime outcomes applied additional benefits to successful treatment of postnatal depression. The changes in strategy for each decision perspective may appear small but would have substantial clinical implications resulting in treatment provision for a larger proportion of postnatally depressed women.

This evidence can be used to update current NICE (2018) recommendations which do not account for lifetime effects in children or establish cost-effectiveness across a full range of EPDS thresholds. Given the large treatment population and high level of decision uncertainty, there is a significant potential value in future research, particularly if the research further clarifies the sensitivity and specificity for different EPDS thresholds.

Chapter 7: Discussion

7.1 Chapter Summary

The overarching aim of the research described in this thesis was to address methods in the economic evaluation of childhood health technologies where analyses could be limited due to the unavailability of evidence across the lifetime. The specific aims were to develop and describe a methodology for indirectly estimating the lifetime effects of early childhood circumstances by linking results across two empirical analyses and to determine whether lifetime estimates influence the cost-effectiveness results in an applied, UK based, economic evaluation. The purpose of this chapter is to draw together the outcomes of the research, discuss in more detail aspects of the results and methodological approach, and highlight how the methodology could be applied in other research scenarios.

7.2 Research Findings

7.2.1 Key Findings

The methodological approach applied in this research provides additional information to decision makers by indirectly estimating the lifetime effects of children exposed to postnatal depression and including these estimates in an applied economic evaluation assessing cost-effectiveness of screening strategies for postnatal depression, extending and updating a previous NICE (2018) analysis.

The overall outcomes of the applied research demonstrated that indirectly estimated lifetime effects *can* influence cost-effectiveness results.

The significance of use of estimated lifetime effects in decision making was shown by comparing results from analyses conducted with and without children's lifetime outcomes. Postnatal depression exposure was predicted to reduce children's lifetime Quality Adjusted Life Years, increase healthcare and crime costs, and generate fewer monetary returns in education and employment. Cost-effectiveness results changed when including children's lifetime effects, leading to the recommendation of a less specific threshold for the Edinburgh Postnatal Depression Scale screening instrument. This change in strategy would have potential clinical implications, providing treatment to a larger proportion of postnatally depressed women

In further detail, when *excluding* children's lifetime effects, the cost-effective screening strategy for postnatal depression was the administration of Whooley questions followed by the Edinburgh Postnatal Depression Scale (EPDS), for all positive initial cases, at a highly specific threshold of 16. Similar strategies but with less specific EPDS thresholds were cost-effective when *including* children's lifetime health (=14), and health & education (=10) outcomes.

The development of methodologies for indirect estimation was the focus of the staged empirical research process described in chapters 3, 4, and 5.

Chapter three demonstrated methodologies appropriate for the first stage of indirect estimation. A primary empirical analysis estimated the effects of early life circumstance (PND symptoms) on *intermediate* lifespan development endpoints identifying that postnatal depression was detrimentally associated with children's cognitive and socioemotional development up to age 11.

The second stage of indirect estimation was informed by a primary analysis undertaken in chapter five. A mathematical model was developed using data from the 1970 British Cohort Study to predict *lifetime decision endpoints* from measures of cognitive and socioemotional development. The full indirect estimation process was completed by entering the incremental effects from chapter three as input parameters in the mathematical model. The output of the model estimated the lifetime effects of children exposed to symptoms of postnatal depression, identifying a strong association between cognitive and socioemotional development at age 10 and later life outcomes relevant as decision endpoints for economic evaluation. This model delivers a potential method for indirectly predicting lifetime outcomes relevant for economic evaluation and UK policy making, thus fulfilling a key research aim in this thesis and filling a gap in the existing evidence base.

7.2.2 Chapter Specific Findings

Individual chapters in this thesis had findings specific to the focus of that chapter and may not be apparent from the report of overall findings. This section summarises these findings and identifies their additional contributions to scientific knowledge.

The literature review in chapter two discussed appropriate methods for economic evaluation in the UK, focusing on the relevant perspectives, decision endpoints and decision rules when the costs and benefits of health technologies occur across sectors beyond health. Whilst the incorporation of cross-sectoral costs within the context of economic evaluation is not new (having been previously applied to the public health literature by Claxton et al. (2010)) no previous studies were identified that applied this theory to early intervention and child development. The review may be useful for

other researchers conducting economic evaluation of early interventions as it defines decision endpoints and type of analyses appropriate when informing UK health policy.

Chapter three adds to existing knowledge by providing evidence on the effects of postnatal *and* subsequent maternal depression in a large longitudinal birth cohort (n=14,000). Previous research conducted by Bell (2014) estimated the effects of maternal depression and its treatment on the development of social & emotional skills and the vocabulary of children up to the age of seven. The research in chapter three extends findings from Bell (2014) by identifying effects (i) up to the age of eleven, (ii) on global measures of cognitive development (rather than just vocabulary); (iii) in models with different empirical specifications. No other literature was identified that estimated the effects of persistent maternal depression for samples sizes of equivalent scale and over a time horizon of similar duration.

The aim of the scoping review reported in chapter four was to add to scientific knowledge by identifying a mathematical model predicting decision endpoints relevant for economic evaluation from earlier measures of child development. Whilst the review was not able to identify any studies which specifically estimated lifetime QALYs and healthcare costs, the null finding is useful as it highlighted a gap in the existing evidence and directed the objectives of the further research in the thesis. The additional synthesis performed for the expanded inclusion criteria identified several characteristics and assumptions aiding the design of the mathematical model in chapter five. These conclusions might be similarly used to inform the design of models elsewhere.

The mathematical model described in chapter five predicted lifetime QALYs, healthcare costs, monetary returns to the education sector (including economic productivity), and monetary costs to the crime sector from measures of cognitive and socioemotional development at age ten. This model delivers a method for extrapolating intermediate measures of child development into lifetime outcomes relevant for economic evaluation and UK policy making, thus filling a gap in existing evidence identified in chapter four. Whilst the model is associated with several limitations, it advances scientific knowledge in the absence of other relevant model and could be applied by other researchers when estimating lifetime effects of other childhood health technologies.

The results of the economic analysis in chapter six provides additional evidence that could be used to update existing policy recommendations regarding screening for postnatal depression. Current economic evaluations of screening strategies for postnatal depression do not consider the lifetime effects in children who are exposed to symptoms.

7.3 Discussion of Research Findings

7.3.1 Introduction

The following section elaborates on the key findings, beginning with a discussion of policy implications drawn from the results of the applied research. The section continues by describing the general importance of including children's lifetime effects in economic evaluation, focussing on two methodological choices which affected their relative influence on cost-effectiveness results: the decision perspective adopted for economic evaluation, and the discount rate applied to costs and benefits. The section concludes by considering whether it would be cost-effective to conduct future research to reduce uncertainty in lifetime effects estimates.

7.3.2 Policy Implications

The major policy implications of the applied research derive from the results of the economic evaluation in chapter six. These suggest that (dual) population screening for postnatal depression would be cost-effective for the NHS if GPs were to administer the Whooley questions to all postnatal women during their first contact with primary care after childbirth and follow up all positive cases with a second screen using the EPDS.

The results of the economic evaluation can be used as evidence by NICE when updating recommendations regarding population screening for postnatal depression. The current NICE guidelines are primarily informed through the results of a NICE (2018) decision model. By extending the structure of the NICE (2018) decision model, research in this thesis demonstrates the impact of children's lifetime outcomes on cost-effectiveness results. Namely, this is to reduce the recommended diagnostic threshold for the EPDS from 16 to 14. The results suggest more sensitive postnatal depression screening strategies become cost-effective if decision makers extend their perspective to account for both maternal *and* child outcomes.

When making policy recommendations, decision makers ought to interpret evidence from economic evaluations by considering both the validity of modelling assumptions and the applicability of results to local clinical practice (Heyland et al., 1999). One of the key implications of screening for postnatal depression at the recommended EPDS threshold (14) is to increase the uptake of treatment in primary care. The potential for screening to increase treatment uptake is contingent on two clinical assumptions made within the original NICE (2018) model, where GPs

are expected to: (i) use screening tools to obtain a postnatal depression diagnosis; and (ii) follow up on positive diagnoses with an appropriate treatment.

Evidence from the broader literature on depression screening in the *general* population raises uncertainty regarding the above assumptions: In a systematic review, Gilbody et al. (2001) identified nine randomised studies which administered psychiatric screening tests to patients in primary care. Each study assigned patients into either an intervention group where test results were fed back to GPs, or a control group where results were not fed back. Feedback did not increase the recognition of depression or anxiety, lead to an increase in treatment uptake, or improve patient outcomes (Gilbody et al., 2001).

There are a variety of reasons why GPs may fail to act upon the results of psychiatric screens. For example, Dorwick et al. (2009) conducted qualitative interviews in a sample of GPs, where the majority viewed their own clinical judgement, rather than screening, as the gold standard for obtaining a diagnosis. This could explain findings by Kendrick et al. (2009) who identified significant associations between anti-depressant prescription/psychological referral rates and scores on depression screening questionnaires, but also found other factors such as patient age, depression history and comorbidities to independently influence diagnoses. In addition, GPs may be restricted by the resources available to treat depression. GPs consistently identify long waiting times, poor access to treatment, lack of available consultation time, and a lack of training as reasons for not offering psychological therapy to depressed patients (Toner et al., 2010).

The evidence from the wider depression literature leaves policy makers with a difficult decision regarding guidelines for postnatal depression screening. The results from this thesis suggests population screening can be cost-effective if GPs are willing/able to administer tests and act upon their results. However, if screening fails to change treatment practices then, as stated by Gilbody et al. (2001), its implementation is a "costly and bureaucratic exercise". If recommending screening, policy makers should ensure appropriate treatments are available. This may require the expansion of local services and additional resources for psychological therapies, which are typically the preferred treatment option for postnatal women (Milgrom et al., 2015).

7.3.3 Influence and Importance of Lifetime Estimates in Economic Evaluation of Childhood Health Technologies

A key requirement for economic evaluation is to account for all the *important* costs and benefits of a technology (Drummond et al., 2015) and/or, to include all *relevant* evidence in an analysis (Sculpher

et al., 2006). Some economic evaluations of childhood health technologies do not account for lifetime effects despite evidence suggesting the potential significance of their inclusion. For example, early childhood visual impairments are associated with cognitive (Bruce et al., 2017) and fine motor (Webber et al, 2008) development but current polices assessing the cost-effectiveness of vision screening in pre-school children do not account for subsequent lifetime outcomes (Carlton et al., 2008). There is a well-established causal association between hearing and development (Moeller et al., 2000) but Bond et al. (2009) did not include lifetime effects in an economic evaluation of childhood cochlear implantation. Meanwhile, Wright et al. (2015) describes how parenting interventions benefit socio-emotional outcomes for children with conduct disorder but could not address these in an economic evaluation because of insufficient evidence regarding lifetime effects. And finally, the applied example in this thesis expanded on a NICE (2018) economic evaluation assessing the cost effectiveness of screening for post-natal depression which did not include lifespan development effects in children exposed to symptoms, describing these as "intangible".

Whether evidence is relevant or important arguably depends on its relative influence on cost-effectiveness results. If results remain unchanged irrespective of the inclusion/exclusion of evidence, then it could be deemed irrelevant/unimportant. The results of the applied economic evaluation in this thesis demonstrated the influence and consequently the importance/relevance of including lifetime estimates as the cost-effective screening strategy for postnatal depression changed for analyses which included/excluded children's lifetime effects. The results of this research therefore *do not* justify the exclusion of lifetime effects from economic evaluation.

The consequence of excluding relevant evidence from economic evaluation, according to Sculpher et al. (2006), is to "risk providing a partial analysis with potentially misleading results". The exclusion of lifetime effects in individual evaluations could risk misleading individual healthcare allocation decisions. The routine exclusion of lifetime effects from all economic evaluation could introduce a general bias into decision making. In general, evidence from this thesis suggested that interventions which improve child development outcomes are likely to translate into more lifetime benefits and less lifetime costs across all sectors. Childhood health technologies which benefit cognitive and socioemotional development are, therefore, predicted to have a net benefit on lifetime outcomes. If these net benefits are consistently excluded from economic evaluation, then potentially fewer childhood health technologies might be considered cost-effective than perhaps should be. Research advancing methods for estimating lifetime effects, such as the research conducted in this thesis, is important as it addresses a *potential* bias in child health allocation decisions.

7.3.4 Perspectives for Child Health Economic Evaluation

The applied research in this thesis suggests the influence of lifetime effects on policy decisions may depend on the analytical perspective adopted. Chapter six identified different cost-effective strategies for postnatal depression when conducting economic evaluation from health centric and cross-sectoral perspectives. The identification of different cost-effective strategies might lead to questions regarding the most appropriate decision perspective. Whether this is a relevant question depends on the role assigned to economic evaluation.

Economic evaluation is sometimes viewed as a tool to prescribe decisions by establishing whether a health technology/social policy provides an overall social benefit by improving social welfare (Drummond et al., 2015). If economic evaluation is prescriptive then it might make most sense to account for all the costs and benefits of a health technology, irrespective of where they fall rather than just accounting for health benefits. Therefore, some might consider the cross-sectoral perspective as most appropriate in the economic evaluation of early childhood interventions as the perspective can account for a range of widespread costs and consequences.

However, economic evaluation has tended to adopt a more modest role in the health sector, being typically used as a tool to inform (rather than prescribe) decisions (Drummond et al., 2015), (Claxton et al., 2007). In this case, the question as to the appropriate perspective ought not to be a theoretical choice made by the analyst, but be a choice based on the pragmatic restrictions of the decision maker whom the evaluation is intended to inform. The applied example presents two different analyses according to different decision perspectives which represent pragmatic conditions in the UK. One assumes a decision maker has the objective of only maximising health with respect to the healthcare budget; and the other assumes the objective is to maximise benefits across multiple sectors subject to the costs incurred in those sectors.

The benefit of adopting multiple perspectives is to demonstrate differences in cost-effectiveness results and any trade-offs that decision makers might face. In chapter six, the cross-sectoral perspective recommended the Whooley questions followed by the EPDS 12 strategy (rather than the EPDS 14 strategy for health centric perspective) as cost effective when screening for postnatal depression, *if* the decision maker is willing to trade off additional costs incurred in the health sector with monetary benefits occurring in the education sector. This trade-off decision must also account for the relative consumption in health relative to consumption in education, which was represented as the v/k ratios (v=consumption value of health, k=monetary value of each additional QALY). The results remained robust for the cross-sectoral perspective for all v/k ratios up to a limit where

consumption in health is assumed to be four times the value of consumption in education. Whilst this would seem to be a reasonable trade-off, the ultimate decision is at the discretion of policy makers – if the policy maker's objective is exclusively health, subject to money spent in the health sector then presumably no trade-off would be acceptable.

It might be more relevant to question whether the health sector should necessarily incur the costs of childhood "health" technologies, or whether this burden ought to fall on the sector accruing most benefit (i.e. education). For the application in this thesis, it would make most sense for the health sector to fund screening for maternal postnatal depression as the largest benefits are expected to be accrued in the health sector in terms of increased *maternal* QALYs and reduced *maternal* healthcare costs. For other interventions, which exclusively benefit children, it may be that the education sector accrues the largest net benefit: As described in chapter five, the relative magnitude of benefits from unit changes in measures of cognitive and socioemotional development was much larger for the education sector than for the health sector. This could result in technologies which are not justified through later QALYs and decreased healthcare costs but are through increased returns in education (i.e. increased human capital).

However, in the example in this thesis, it is not certain whether the benefits of child development are actually larger in education when compared with the health sector. Larger magnitudes may have been observed in chapter five because more data was available regarding children's lifetime returns to education. Several limitations in the analysis of health outcomes were noted, most importantly the ending of the observed study time horizon at age 42 when most health effects likely to occur after the study end point. It is possible the analysis underestimated the increased QALYs and reduced healthcare costs associated with improved child development outcomes. On the other hand, a lifetime horizon may have identified increased healthcare costs as a result of extensions to life expectancy. As described in chapter five it would be useful to estimate a model predicting lifetime QALYs/healthcare costs from measures of child development using appropriate questionnaires (i.e. items to identify health related utility e.g. EQ-5D/and resource use questionnaires, over an extended time horizon).

7.3.5 Discount Rate Applied to Lifetime Decision Endpoints

The results of chapter five (Section 5.5) demonstrated the substantial difference discounting can have on the magnitude of lifetime predictions per unit change in cognitive and socio-emotional development measures. Further, results from chapter six identified different cost-effectiveness

results for scenarios estimating lifetime effects at different discount rates (Section 6.4/6.5). As discounting can have fundamental effects on outcomes it is discussed in more detail in this section.

As the lifetime effects of early life circumstances are expected to occur far into the future, different discount rates can lead to very different values assigned to costs and benefits. For example, a cost equal to £100,000 which occurs fifty years into the future would have a present-day value equal to £47,500 discounted at a rate of 1.5%, or equal to roughly £18,000 if discounted at 3.5%.

The appropriate rate to discount costs and benefits is uncertain as it depends on factors which might change over time. Firstly, the calculation of discount rates depends on the relative discount rate of consumption, usually calculated by economists using the Ramsey rule which accounts for both society's time preferences (i.e. prefer consumption now rather than later) and for changes in the value of consumption over time (i.e. decreases in value of money as a result of increases in economic productivity) (Arrow et al., 2014). Neither time preference nor the value of consumption is necessarily constant, for example, political/environmental instability might result in society placing more weight on current utility if prosperity in the future is uncertain. Meanwhile, a myriad of factors could influence economic productivity, e.g. stock market crashes, technological advancements, political conflicts, environmental disasters etc.

In addition to uncertainty regarding the appropriate discount rate for consumption, changes to the consumption value for health, the budget for healthcare, and the cost-effectiveness threshold can influence specific discount rates applied to health outcomes (Drummond et al., 2015). There is also growing consensus in the literature that health should be discounted at a lower rate than healthcare costs (Attema et al., 2018). Claxton et al. (2011) advocates lower discount rates for health if it is assumed the budget for healthcare is not fixed and that consumption in health grows relative to consumption in other sectors. Claxton et al. (2011) suggests this a likely scenario as "the health of the population is expected to grow more slowly than its consumption of goods and services" meaning "welfare gain from better health will increase relative to welfare gain from increases in consumption". The application of different discount rates on costs and QALYs can have implications on cost-effectiveness results, as was demonstrated for the health centric decision perspective in chapter six (Appendix 6.4G).

In the face of uncertain discount rates likely to influence cost-effectiveness results, an analyst's best response might be to present the results for several scenario analyses applying different assumptions regarding the discount rate. Currently, the Treasury (2018) Green Book suggests a 3.5% discount rate, which is reflected in NICE (2013) guidelines. The research in this analysis followed NICE (2013) guidelines for healthcare costs/monetary returns (3.5%) but also identified cost-effectiveness

for scenarios with QALYs, healthcare costs and consumption costs discounted at 1.5%, and different discounting costs (3.5%) and QALYs (1.5%). These scenarios could be extended by investigating a larger range of discount rates. They might also be extended by applying different discount rates according to when outcomes occur, for instance Arrow et al. (2014) recommend the application of declining discount rates over time.

7.3.6 Cost-Effectiveness of Future Research

The results from chapter six suggest the potential cost-effectiveness of future research depends both on the perspective adopted for evaluation, and the value of the cost-effectiveness threshold. The Expected Value of Perfect Information (EVPI) was much larger for the cross-sectoral than the health centric decision-making perspective. Meanwhile, the Expected Value of Partial Perfect Information (EVPPI) for the child parameters was equal to zero at a cost-effectiveness threshold of £20,000, but equal to almost £50 million at a cost-effectiveness threshold equal to £30,000.

Future research could attempt to reduce the uncertainty associated with indirect lifetime estimates. The preferred method of reducing uncertainty is to include multiple sources of evidence from secondary literature sources and methods for evidence synthesis are discussed in more detail in section 7.4.3. Uncertainty could also be reduced by taking forward research priorities identified at the end of each empirical chapter. These included the estimation of a multivariate growth curve model to identify the effects of postnatal depression on joint outcomes for cognitive and socioemotional development (chapter three) and a path model to account for adulthood variables when predicting lifetime outcomes from intermediate measures of child development (chapter five).

7.4 Discussion of Methods for Estimating Lifetime Effects

7.4.1 Overview of Methodological Approach in this thesis

The methodological approach applied in this thesis *indirectly* estimated the incremental mean lifetime QALYs, healthcare costs, monetary costs/returns in education and crime sectors for children exposed to symptoms of postnatal depression vs. children not exposed, as depicted in Figure 7.1. Initially, the incremental effects of exposure to postnatal depression (vs. non-exposure) were identified on intermediate outcomes for children's cognitive (General Cognitive Ability (GCA)) and socioemotional development (Strength and Difficulties Questionnaire (SDQ)) in chapter three. Growth curve modelling was used to observe the association between maternal postnatal

depression and GCA/SDQ scores at four separate intervals in a time series database allowing mean exposure effects and changes in exposure effect to be identified between the ages of three and eleven.

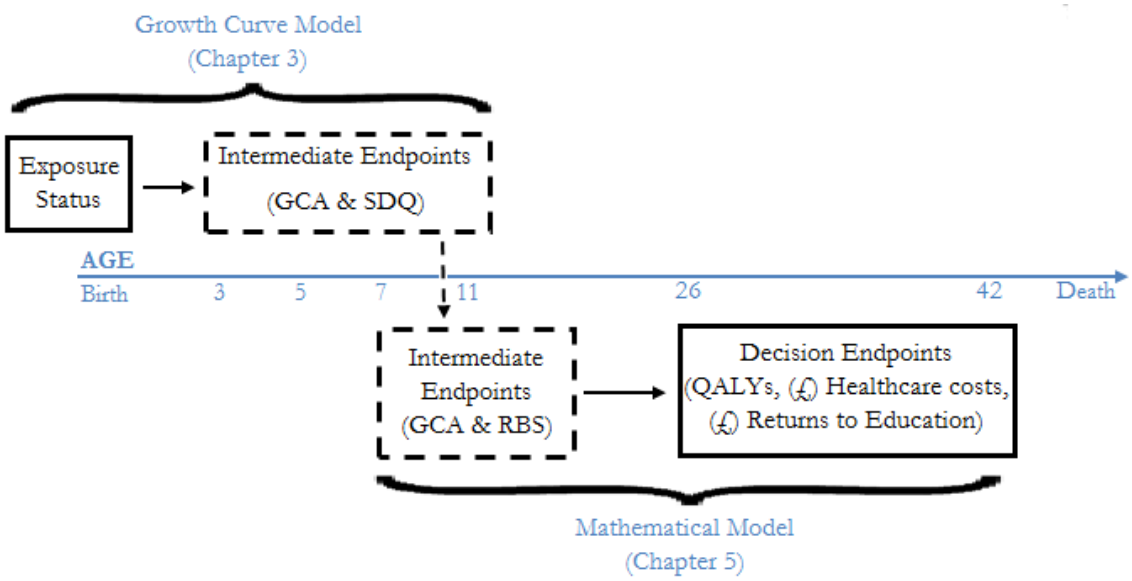


Figure 7.1: Depiction of the methodological approach for indirect estimation.

Chapter five addressed the second stage of indirect estimation, which required effects of postnatal depression on intermediate outcomes to be translated to effects on final decision endpoints. Mathematical models were used to predict the effect of a unit change in measures of cognitive (GCA) and socioemotional development (Rutter Behaviour Scale (RBS)) on lifetime QALYs, healthcare costs, and monetary returns to education/crime sectors. Mean incremental lifetime effects in children exposed to postnatal depression were predicted by linking the results from the two studies. Coefficients attached to the GCA/RBS variables in the mathematical model (chapter five) were multiplied by the corresponding incremental effect size identified for GCA/SDQ outcomes in the growth curve models (chapter three).

7.4.2 Validity of Methodological Approach

7.4.2.1 Validity of Indirect Estimation

An implicit assumption of indirect estimation is that the effects of an early life circumstance on lifetime endpoints can be predicted without the need for direct observation. Instead, lifetime effects are identified by linking the results of two direct studies which (a) observe the relationship between early life circumstances and intermediate outcomes and, (b) observe the relationship between intermediate outcomes and lifetime endpoints. As indicated in chapter one (Section 1.2.5) questions addressing the validity of this assumption, and the general appropriateness of indirect estimation, can be informed by considering the surrogate endpoints literature.

A common application of indirect estimation in health economic evaluation is the use of *surrogate endpoints,* which act as substitutes (or intermediates) for final clinical or decision endpoints (Hawkins et al., 2012). Surrogate endpoints are often referred to as biomarkers and might include measures of biological process, pathogenic process, or pharmacological response to a therapeutic intervention (Aronson, 2005). As a biomarker, surrogate endpoints are assumed to occur on the treatment pathway before the final endpoints, Figure 7.2. Examples of validated surrogate endpoints include HIV infection to predict final clinical endpoints related to AIDs (Fleming and Powers, 2012) and blood pressure as a biomarker for later cardiovascular events (Aronson, 2005).

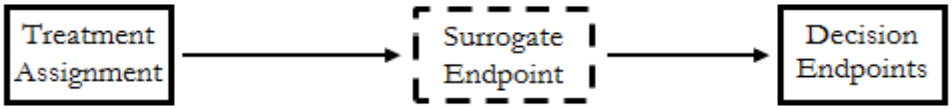


Figure 7.2: Illustrates the path from treatment to decision endpoint via effects on a surrogate endpoint.

According to Fleming and DeMets (1996), a common misconception is to assume validity if there are associations observed between treatment assignment and the (proposed) surrogate endpoints, and between the surrogate and final endpoints. For a surrogate outcome to be a true substitute for a decision or clinical endpoint, it must predict the full effect of an intervention on the final endpoints and not merely be correlated (Fleming and DeMets, 1996). Four criteria proposed by Prentice (1989) have been widely adopted to assess the validity of surrogate endpoints. As demonstrated by the inserted parenthesis, the Prentice (1989) criteria can equally be applied to assess the validity of

cognitive and socioemotional development measures as intermediate outcomes during indirect estimation of lifetime endpoint:

1. Treatment (an early life circumstance) has a significant impact on the surrogate (intermediate) endpoint(s).
2. The surrogate (intermediate) endpoint(s) has a significant impact on the true endpoint.
3. Treatment (an early life circumstance) has a significant impact on the true endpoint.
4. The full effect of treatment (an early life circumstance) upon the true endpoint is captured by the surrogate (intermediate) endpoint(s).

The research in this thesis immediately fulfils the first two requirements. Significant associations were identified between postnatal depression and cognitive and socioemotional outcomes (chapter three), and between cognitive and socioemotional outcomes and lifetime decision endpoints (chapter five). In contrast, it is not possible to formally validate the third and fourth criteria as both require direct evidence of the effects of postnatal depression on lifetime decision endpoints. The reason for using indirect estimation here was specifically due to the unavailability of direct evidence.

Whilst formal validation is not possible, justification for the third criteria was previously summarised in the literature review in chapter two, which identified an effect pathway between early life circumstances and lifetime decision endpoints. The review suggested early life circumstances, such as postnatal depression, can influence the *individual* ⟷ *context* relation (the mechanism driving lifespan development) which necessarily determine an individual's function, behaviour and capabilities across the remainder of their lifespan.

Whilst not specifically related to postnatal depression, there is some direct evidence from randomised controlled trials confirming the pathway of effect between early life technologies/interventions and adulthood decision endpoints. For example, Currie (2001) reviews the Head Start preschool programme, a group of educational interventions that began in 1965 and aimed to boost the development of underprivileged children in the USA. Three interventions were found to have positive effects on adulthood outcomes in long term randomised controlled trials when compared with control. The Perry school project reported beneficial effects on academic attainment and earnings, and negative effects on crime participant and welfare use (Currie, 2001), (Heckman et al., 2010); the Carolina Abecedarian Project treatment group achieved higher academic test scores, were more likely to be in further education at age 21 (Currie, 2001), and had more years of schooling than the control group at age 30 (Campbell et al., 2012) and the Early Training Project intervention group achieved (non-significant) increases in academic achievement and high school graduation (Currie, 2001).

Despite the evidence supporting the existence of a significant effect pathway (criteria three), it is not possible to validate the fourth criteria and establish whether the *full effects* of early life circumstances on children's final endpoint are captured on intermediate measures of child development. Other effect pathways might exist between maternal depression and children's lifetime outcomes. For example, postnatal depression could feasibly influence family socioeconomic status (SES); family SES might determine children's SES in later adulthood separate from the effects of family SES on child development (i.e. poor children are likely to remain poor regardless of their development status); and adulthood SES might be associated with relevant adulthood outcomes e.g. heath, earnings, education, crime etc. Figure 7.3.

Whilst multiple effect pathways may exist between early life circumstances and lifetime decision endpoints, it would seem likely that these occur predominantly in the same direction as the indirect estimates. That is, it is difficult to think of legitimate reasons why maternal depression would have a *positive* impact on children's later life outcomes. So, while it is not possible to validate the fourth criteria, it might be reasonable to suggest the effects of maternal depression on children's lifetime outcomes at least match those identified in this thesis. Even if the indirect methods underestimate the true magnitude of association between early life circumstances and adulthood decision endpoints, their inclusion in economic evaluation still represents an improvement on the initial decision-making conditions. It is better to include some of the overall effect rather than exclude it from the analysis completely.

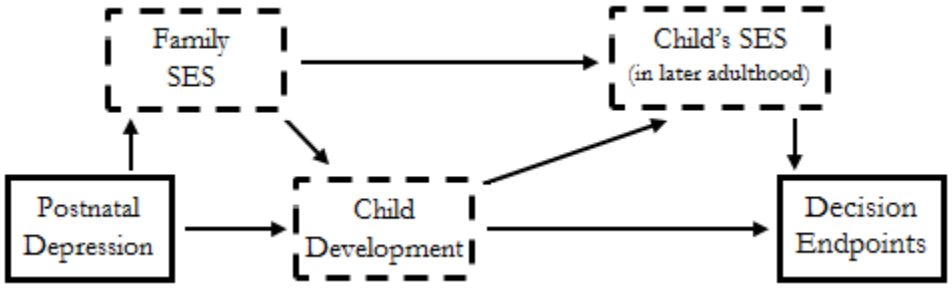


Figure 7.3: A *hypothetical* effect pathway between postnatal depression and children's lifetime decision endpoints. Effects may occur through the indirect pathway which includes effects to child development. Different effect pathways may also exist, for example through the impact of postnatal depression on family, and later child socioeconomic status (SES).

7.4.2.2 Additional Structural Uncertainties within Indirect Estimation

In addition to assuming the appropriateness of the general pathway for indirect estimation, two potentially questionable assumptions were made within the process. Firstly, the data available from the empirical studies used slightly different measures to assess children's socioemotional development, where the Strengths and Difficulties Questionnaire was used in chapter three, and the Rutter Behavioural Scale used in chapter five. Secondly, it was assumed the incremental effects of postnatal depression on child development remained constant beyond the observed time horizon in chapter three (after age eleven).

Uncertain assumptions in analytical methods and scientific judgements are termed structural uncertainties by Bojke et al. (2009). The most common method of addressing structural uncertainties in economic evaluation is to conduct a deterministic sensitivity analysis and present results across scenarios which differ in terms of the structural assumptions made (Bojke et al., 2009). When assessing the cost-effectiveness of screening for postnatal depression, the research in chapter six conducted deterministic sensitivity analysis for the two structural uncertainties identified above. Cost-effectiveness was established for scenarios where (a) indirect estimation was conducted with an intermediate measure for socioemotional development which only included categories appearing on both SDQ and RBS scales (Appendix 6.5), and (b) increasing/decreasing development effect sizes were assumed when estimating lifetime outcomes by multiplying the estimated effects by a proportion of 0.75 or 1.25 (Appendix 6.4). Neither structural assumption altered the key findings in this research, this being the potential for lifetime effects to influence the cost-effectiveness result.

Whilst not affecting the key findings, different structural assumptions did influence the specific screening strategy recommended within the deterministic scenarios (Appendix 6.4). In this case, presenting results across several scenarios may not appropriately inform decision makers as they may not be able to determine the relative importance of each structural uncertainty. This analysis could have been extended by applying a technique called model averaging to identify a single cost-effectiveness result whilst accounting for results across multiple scenarios (Bojke et al., 2009). Model averaging requires weights to be applied to each uncertain scenario relative to the likelihood of that scenario being the correct assumption (Bojke et al, 2009). Cost-effectiveness across all scenarios could be established by calculating incremental cost-effectiveness ratios (ICERs) across weighted QALYs and healthcare costs.

7.4.2.3 Causality in Observational Study Designs

The previous sections considered the validity of the indirect estimation procedure. There may also be concerns regarding the validity of the two pieces of empirical evidence used to inform the indirect estimation of lifetime effects in this thesis, as both chapters three and five utilised data form observational birth cohorts. As stated by Kaplan (2018), observational studies identify correlations, but do not prove causation.

It could be that the association between maternal depression and child development, identified in chapter three, occurred due to factors unaccounted for in the growth curve analysis. For example, children might have been exposed to antenatal depression symptoms which also exhibit detrimental associations with child development (Kingston and Tough, 2014). However, this may be unlikely given that several studies have identified associations for *both* antenatal and postnatal depression within the *same* model (Kingston and Tough, 2014). Meanwhile, depression genes could be passed from mothers subsequently manifesting as behavioural/emotional problems during childhood, though this would not explain the relationship between maternal depression and cognitive development outcomes.

There are several reasons why a causal relationship might be assumed for the associations identified in chapter three. Firstly, there is abundant theoretical literature assigning an effect pathway between maternal depression symptoms and child development through the maternal-child attachment relationship (Bowlby, 1978). Secondly, the theoretical literature is supported by evidence from experimental studies where depression interventions are found to decrease symptoms, improve the maternal-child attachment relationship and positively influence early child development (Cuijpers et al., 2015). Thirdly, associations between maternal depression and child development have been consistently identified in other observational study designs (Sanger et al., 2015). Finally, causality is supported by associations which exhibit biological gradients i.e. dose-dependencies (Fedak et al., 2015), and the growth curve models identified a strong accumulative effect of maternal depression symptoms on both child development outcomes.

A similar rationale can be used to infer causality for the associations between measures of child development and lifetime outcomes in chapter five. Again, dose dependent responses were identified where each unit increase in child development measures resulted in increases in the magnitude of association with lifetime outcomes. Meanwhile, chapter four identified evidence from several observational studies where consistent associations were identified between child development and different adulthood outcomes including earnings (Cunha and Heckman, 2008),

well-being (Layard et al., 2014), self-rated health (Hertzman et al., 2001), and high school graduation (Cunha et al., 2010). The causal assumption is also supported by the theoretical and empirical evidence used to previously support the validity of indirect estimation (in section 7.4.2.1).

The best way to address validity issues related to causality for both empirical studies might be to combine all the available evidence using methods for evidence synthesis. Appropriate techniques for evidence synthesis during indirect estimation are discussed in the following section.

7.4.3 Evidence Synthesis and Indirect Estimation

As described in section 7.2.3.1, economic evaluations should be informed by all the relevant evidence (Sculpher et al., 2006). The research in this thesis partially addresses this requirement, developing a method for indirect estimation which provides a mechanism to use additional evidence in economic evaluations by linking findings from two empirical studies. To comply fully with the requirements suggested by Sculpher et al. (2006), the methodology for indirect estimation ought to be expanded to include evidence from all relevant sources.

There are likely to be relevant sources of evidence available in the current literature base to more fully inform the first stage of indirect estimation, which requires effects from early life circumstances to be identified on intermediate development outcomes. For example, there are several sources of evidence establishing the association between postnatal depression exposure and measures of cognitive and socioemotional development (for examples see Kingston and Tough (2014) & Sanger et al. (2015)).

Typically, health research synthesises evidence on effect sizes in a meta-analysis, a statistical technique combining results from multiple empirical studies into a single pooled (mean) estimate (Sharpe, 1997). The effects of early life circumstances could potentially be estimated using all the available evidence in separate meta-analyses with pooled effects for cognitive and socioemotional development outcomes.

A key assumption of meta-analyses is homogeneity between the included studies. This means pooled effect sizes should only be calculated across studies measuring the same outcomes, against the same comparator, and in the same population (Sharpe, 1997). The synthesis of evidence across studies evaluating effects on child development outcomes might be problematic due to prevalence of study heterogeneity. Study heterogeneity occurs due to the diverse outcome measures available to assess child development. For example, when estimating effects of exposure to postnatal depression symptoms, systematic reviews by Kingston and Tough (2014) and Sanger et al. (2015) identified ten

different cognitive development measures and eight socioemotional development outcome measures.

The problem of outcome heterogeneity might be addressed in multiple operationalism, a method allowing different types of measures to be combined in a meta-analysis if the heterogeneous outcome measures identify a similar underlying construct. Behavioural and social sciences often rely of multiple operationalism meta-analyses as studies regularly use diverse measures that capture different facets of the same latent variable (Noble Jr, 2006). The technique has been applied by Beck (1998) to estimate a pooled effect for children exposed to symptoms of postnatal depression on cognitive and emotional developmental outcomes occurring throughout infancy and early childhood.

A second source of study heterogeneity may occur if studies identify effects on development outcomes at different ages. Chapter three described how growth curve models were required to account for potentially changing effect sizes which may occur as children develop. Even pooled results in a multi-operationalism meta-analysis may not be appropriate if, for example, one study identifies outcomes at age two, and the other at age eleven.

To avoid heterogeneity, analysts could restrict the meta-analysis by specifying an age range for the development outcome. The appropriate age range is likely to depend on the evidence available in specific research scenarios. Ideally, pooled effects would be obtained on outcomes occurring towards the later stages of child development. This would reduce the potential for changing effect sizes over time, as most developmental change will have occurred by this age. Increasing effectiveness evidence, which could facilitate the conduct of a meta-analysis, is likely to become available due to the additional emphasis being placed on the Early Years literature. For instance, the Born in Bradford Better Start study (BiBBs) was recently commissioned, an experimental birth cohort study aiming to identify effectiveness evidence for over twenty early interventions and follow up children until they are aged ten (Dickerson et al., 2016).

However, in some cases, it may be that too few long terms studies are available to sufficiently inform a meta-analysis across child development outcomes occurring in mid childhood. Most evidence, particularly robust trial-based evidence, is likely to be available across studies with shorter time horizons. For instance; the meta-analysis by Beck (1998) identified eight studies with outcomes collected at ages ranging from one to four, but only two studies with outcomes occurring after the age of four. Analysts may need to trade off the length of the time horizon with the number of available studies.

If indirect estimation were to be informed through a meta-analysis conducted during *early* childhood, some method would be required to predict the size of the pooled effect at age 10 (the age corresponding with the developmental input parameters for the mathematical model in chapter five). As is described in chapter three, statistical methods such as growth curve models can utilise time series data to establish whether mean effect sizes on developmental outcomes were likely to change over time. Evidence from time series analyses could be used to predict any changes expected to occur in the pooled effect between early childhood and age 10 as indicated in the following hypothetical example: If a pooled effect size equal to X was identified on cognitive development outcomes at age three and a time series analysis identified an effect size on cognitive outcomes at age three equal to Y, which changed to an effect size equal to Y/A at age ten; then a pooled effect size for cognitive development outcomes at age ten might be estimated as X/A, assuming the pooled effect changed proportionately to the effect in the long term study. This approach to evidence synthesis is used by the Washington State Institute for Public Policy when establishing the cost-benefit of several early childhood health and social policies (WSIPP, 2017).

The evidence synthesis framework could be further expanded by using different mathematical models to translate intermediate pooled effects to final decision endpoints in the second stage of indirect estimation. Whilst no other models estimating lifetime QALYs and healthcare costs were identified in chapter four, these may become available in the future. It would not be appropriate to pool estimates from different mathematical models in a meta-analysis because models are usually heterogeneous applying a diverse set of structural assumptions. Instead of conducting a meta-analysis, separate lifetime estimates could be obtained for each individual model and used to inform different scenario analyses within an economic evaluation. This method might be extended further by obtaining a single estimated effect size through model averaging, as was described when discussing structural uncertainties in the previous section. The full evidence synthesis framework suggested for indirect estimation is illustrated in Figure 7.4.

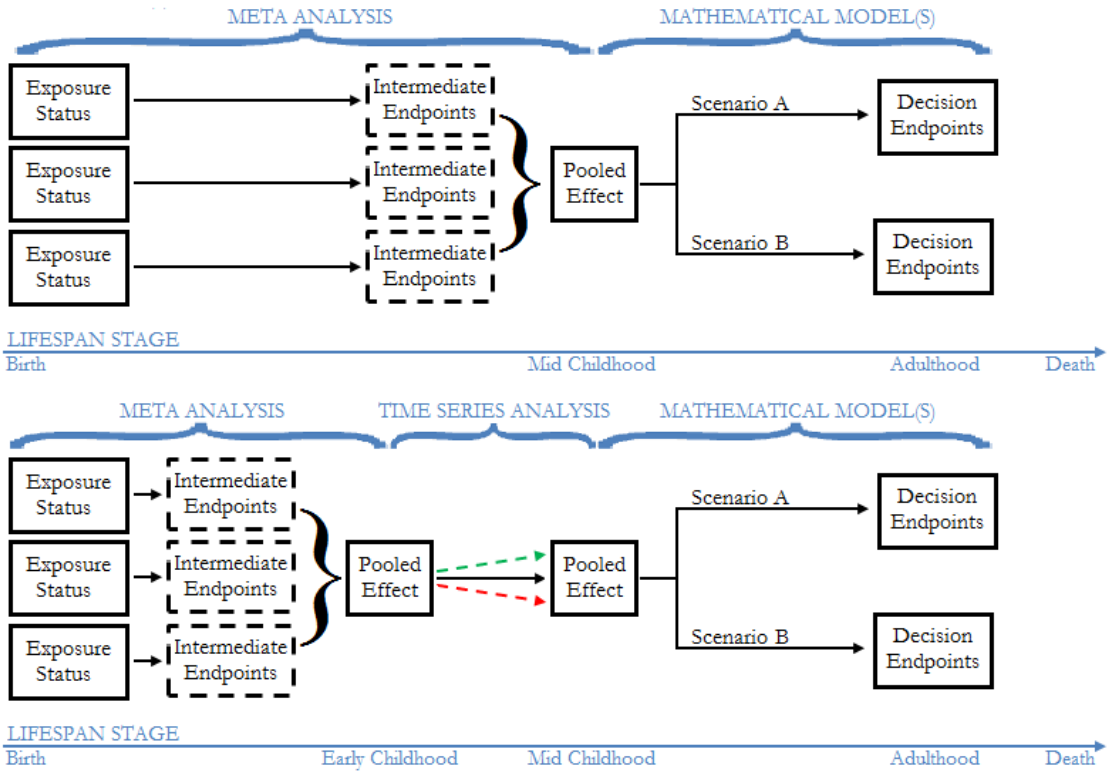


Figure 7.4: Illustrates an evidence synthesis framework that could be applied to estimate the lifetime effects associated with children exposed to postnatal depression symptoms (or other childhood diseases/exposures/health technologies). *Above* (most desirable)*:* Assumes evidence is available to conduct a meta-analysis across intermediate development outcomes in mid-childhood; pooled effects in mid-childhood are extrapolated to adulthood decision endpoints in different scenarios using separate mathematical models.
Below: Assumes evidence is only available to conduct a meta-analysis across development outcomes in early childhood; pooled effects during early childhood are adjusted based on the results of a time series analysis to obtain pooled effects during mid-childhood; adjusted pooled effects in mid-childhood are extrapolated to adulthood decision endpoints in different scenarios using separate mathematical models.

7.4.4 Alternative Approaches to Acquiring Lifetime Evidence

7.4.4.1 Introduction

The methodology applied in this research is one of several approaches that might be used to identify the lifetime effects associated with early life circumstances. This section considers the range of approaches that could be available to researchers and suggests a ranking of different methods for estimating lifetime effects through direct and indirect estimation. This section might provide a useful guide for other researchers.

7.4.4.2 Direct Estimation

Whilst not available for the applied example in this thesis, usually, the best and most reliable sources of evidence are obtained by *directly* measuring effects on relevant decision endpoints within a trial-based design. Randomised controlled trials (RCTs) are often termed the "gold standard" for direct estimation, as they eliminate several sources of bias through random allocation of participants to different (treatment) groups (Akobeng, 2005). Section 7.4.2.1 identified the outcomes of early education programmes including the Perry preschool programme (Heckman et al., 2010) and the Carolina Abecedarian project (Muennig et al., 2011). Such research demonstrates the possibility of conducting randomised trials over a long-term time horizon and identifying relevant decision outcomes during adulthood. However, probably due to high associated costs, there are few other examples in the literature where direct lifetime evidence is available from long term RCTs.

A more pragmatic approach to obtaining direct long-term evidence might be to conduct trials within cohorts, utilising the multiple randomised controlled trial (cmRCT) design advocated by Relton et al. (2010). The cmRCT recruits participants with a condition into a large observational cohort. Different interventions are randomly offered to the cohort members who are followed up similarly to typical longitudinal studies with repeated outcomes obtained across several data collection sweeps. Treatment effects are identified as in RCTs, by comparing the outcomes in intervention and control groups (Relton et al., 2010). When compared with RCTs, cmRCT designs are associated with reduced costs as multiple RCTs can be conducted within the same cohort. Likely due to their lower associated costs, Relton et al. (2010) state that cmRCTs can collect longer term evidence "as standard", which might make them an appropriate study design when estimating lifetime effects.

In some cases, RCTs and cmRCTs might not be ethical if research questions relate to the effects of diseases or exposures. For instance, in this research it would not have been appropriate to knowingly expose a group of children to harmful symptoms of postnatal depression. Where RCTs are unethical, direct evidence can be obtained from observational study designs which include cohort, case control, and cross-sectional studies. Ethical allocation of disease and exposure groups can be achieved within observational studies as this is assigned in nature and not controlled by researchers. Because randomisation is not possible, evidence from observational studies is often interpreted with caution as biases may be introduced particularly if the study groups are imbalanced for certain characteristics that may affect the final outcomes (Webb et al., 2016).

Birth cohort studies are likely to provide lifespan development researchers with the most useful sources of evidence from observational studies as they obtain repeated evidence on participants as they progress from early childhood into adulthood. The most relevant UK birth cohorts that have currently reached adulthood include the 1958 National Child Development Study and the 1970 British Cohort Study. Additional direct evidence from observational studies will also become available in the future when recent birth cohorts, such as the Millennium Cohort (MCS) and the Born in Bradford (BiB) studies, reach adulthood.

Researchers and decision makers might question the appropriateness of direct estimation given the time horizon required between early life circumstance and the final observed outcome. For example, if direct evidence were collected from the 1958 National Child Development Study the evidence would rely on the relevance of historical data collected 50 years ago. A key consideration in economic evaluation is the external validity of the evidence used to inform the analysis i.e. whether findings for study participants hold true for the target population of the intervention (Sculpher et al., 2006), (Murad et al., 2018). Decision makers may be reluctant to rely on direct evidence from the 1950s given the fundamental changes that have occurred in the UK to technology, science, demographics, politics, economics etc. Lifetime effects estimated through indirect methods might be considered more valid as the procedure links multiple sources of evidence with shorter time horizons. Rather than relying on the effectiveness of interventions assessed in the 1950, indirect estimation could utilise evidence from UK birth cohorts commissioned as recently as 2000 (the MCS) or even 2007 (BiB).

7.4.4.3 Using Indirect Estimation

Indirect estimation might, therefore, be conducted if there is either no direct evidence available or if it is considered the most (externally) valid approach to obtain lifetime effects estimates. The evidence synthesis framework described in the section 7.4.3 (Figure 7.4), which incorporates all the available literature, represents the most desirable method for indirect estimation. Where there is insufficient evidence available to conduct evidence synthesis, researchers could adopt a methodology like that applied in this thesis and estimate indirect effects by linking evidence from two (or more) empirical studies.

Data linkage during indirect estimation need not be approached homogenously by all researchers. For example, researchers should consider the age of children when studies are linked, i.e. the age at which intermediate development measures are obtained. It is suggested that where possible, longer term intermediate outcomes ought to be prioritised. In the approach in this thesis the first stage of

indirect estimation *directly observed* the effects of an early life circumstance on measures of child development. In contrast, the second stage of indirect estimation used a mathematical model to *predict* the effects of development on relevant lifetime outcomes. To reduce uncertainty, it would make sense to maximise the time horizon of the more reliable first stage and minimise the time horizon over which effects are being predicted. For this research, intermediate outcomes were measured for ages 10-11 simply because it was the oldest age for which data was available in the Millennium Cohort Study.

Indirect estimation could also be conducted by using multiple sets of intermediate endpoints to link the results of more than two studies. A study by Hummel et al. (2011), for example, estimated the lifetime effects of early parenting interventions indirectly using three stages where: effects of early intervention were obtained on measures of cognitive and socioemotional development at age three; a mathematical model was used to extrapolate these effects to cognitive and socioemotional development measures at age five; and a second mathematical model was used to extrapolate children's age five outcomes to lifetime decision endpoints at age thirty-four.

Ideally extra modelling stages should not be introduced. Whilst the additional modelling stages might improve the external validity of the estimates by further reducing the time horizon of the required evidence, they simultaneously increase the associated structural uncertainty. Each additional extrapolation requires an added assumption about the continuation of an effect pathway which cannot be formally validated in the absence of direct evidence. However, such methods might be appropriate if evidence on the effects of early life circumstances on measures of child development is only available for early childhood.

Figure 7.5 summarises the discussion in this section, ranking different methods for estimating lifetime effects through direct and indirect estimation which might provide a useful guide for other researchers.

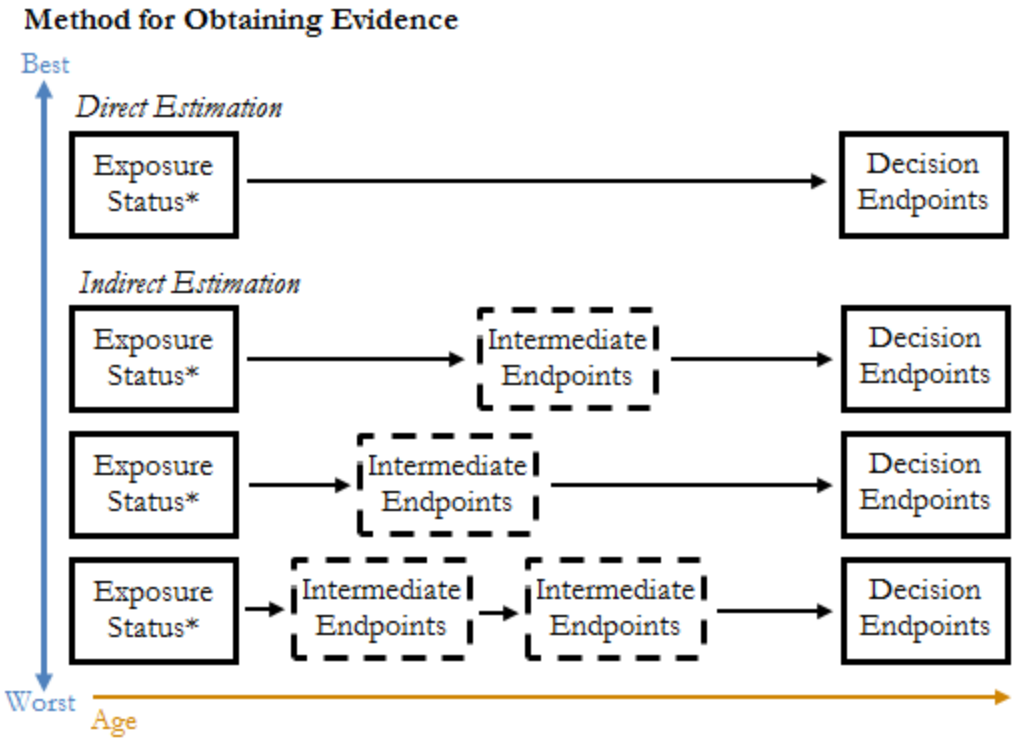


Figure 7.5: Hierarchy of evidence when estimating the effect of early life exposures that affect child development on lifetime decision endpoints. *Can be interchanged with a different early life circumstance according to study design e.g. disease status, treatment status etc.

7.4.4.4 Expert Opinion

If there is insufficient direct and/or indirect empirical evidence, economic evaluation could be informed through expert opinion where specific parameter values are established through estimates provided by expert professionals or researchers. As expert opinion is subjective, it has often been viewed with caution (Leal et al., 2007). Recent research has attempted to establish structured methods to quantify experts' beliefs to increase the reliability of the derived parameters (Soares et al., 2018). If expert elicitation is used to inform economic evaluation, methods are required to identify both a mean value and the likely distribution for the parameter of interests. Soares et al. (2018) describe two methods that can be used to quantify expert opinion, these being: the fixed interval method, where experts are provided with a range of values and assess the probability that the value of a parameter lies between the interval; and the variable interval method where experts are asked to attach values to specific predefined percentiles of the distribution.

There are several reasons why expert opinion might not be appropriate when obtaining estimates for lifetime outcomes. Firstly, expert opinion requires evidence to be obtained from professionals

with enough expertise in the parameter of interest. For instance, whilst GPs might have sufficient patient experience to provide useful and informed beliefs regarding the likely effects of a particular treatment and be in a position to observe effects of certain diseases/exposures/treatments on a child's early development, it is unlikely that they have necessary experience of health economics to understand how these effects might influence lifetime health, education and employment and crime outcomes. Similarly, economists and social researchers may be able to provide opinions on lifetime decision endpoints but might be unlikely to have relevant experience regarding the effects of the early life circumstance on their early health and development.

Secondly, and as described by Soares et al. (2018), expert opinion can be problematic as individuals often find it difficult to quantify their beliefs, particularly if they are not trained in quantitative methods. It would seem particularly difficult in the case of lifetime estimates, as these require complex calculations. For example, experts estimating incremental effects on lifetime earnings might need to first calculate mean lifetime earnings by aggregating individual earnings over the duration of the employment; and then multiply this value by an appropriate proportion according to the expected size of the incremental effect. Analysts should carefully consider whether lifetime estimates informed through expert opinion are a reliable source of information for economic evaluation.

7.4.4.5 Modelling Development within a Decision Analytic Model

Direct/indirect estimation and expert elicitation might all be used to obtain evidence regarding the incremental lifetime effects associated with a specific early life circumstance versus a relevant comparator. Such evidence can be used to estimate parameter values attached to different states in decision analytic models (DAMs). For example, this research obtained and assigned lifetime QALYs, healthcare costs and returns to education for health states where children were categorised as either exposed or not exposed to symptoms of postnatal depression. A similar approach might also be utilised in trial based economic evaluations requiring identification of the lifetime outcomes associated with a specific intervention versus a relevant control (e.g. usual care).

Alternatively, economic evaluations could account for lifetime consequences of early life health technologies by incorporating the process of child development within the structure of a DAM. Typically, DAMs identify the effects of health technologies by assuming that a cohort of hypothetical patients pass through a series of different *health/disease* states; for example, an intervention to treat cancer might be assessed in terms of its effect on patients' progression through states for disease, remission, relapse and death.

Equally, "development" states might be used to capture the effects of early health technologies in terms of how they affect the progression of a hypothetical cohort of children through the development process. When identifying the cost-effective treatment for postnatal depression Ride (2018) theorised a Markov model that contained "development" states, classifying children as either having or not having a cognitive impairment following their exposure to symptoms. The model by Ride (2018) represents a small expansion to the DAM in chapter six as it moves beyond disease states in mothers (depressed/not depressed) and models states during early childhood (cognitive impairment/no cognitive impairment). The model might be expanded upon further by including additional states that extended later into the development process (e.g. cognitive impairment in adolescence/no cognitive impairment in adolescence).

There are two criticisms that might be directed at methodologies attempting to incorporate the progression of development within the structure of a DAM. Firstly, the approach assumes that development can be appropriately specified into specific qualitative states and this might be controversial. For instance, Ride (2018) dichotomises children as having cognitive impairment or not and, therefore, requires a diagnosis for cognitive impairment – this is assigned based on whether children score below an arbitrary threshold on a psychometric test. This dichotomy might be inappropriate in the context of this thesis if it was assumed that exposure to postnatal depression could have detrimental effects across all children. For instance, children initially predisposed as high achievers would have to be severely affected for them to reach a diagnosis that moved them to a state of cognitive impairment. As defined in chapter two, development is change that occurs within *individuals* over time and is likely to occur (at least partially) as a continuous process. It may not make sense to model *continuous* individual changes within *discrete* states in a DAM.

A second criticism relates to the pragmatic restrictions that could make it impossible to estimate a DAM that appropriately captures the process of lifespan development. It might be argued that the framework suggested by Ride (2018) should be expanded into a more complex structure including in this case additional effects of postnatal depression on socio-emotional development. The inclusion of socio-emotional effects would require exponential expansions to the total number of development states included in the model e.g. no impairment, cognitive impairment only, socio-emotional impairment only, both cognitive and socioemotional impairment. A *fully* representative model of lifespan development that could be utilised more generally to assess the cost-effectiveness of wide range of early health technologies might require more additional development states e.g. in physical and environmental domains.

Further complexity might be incorporated in a development DAM given the potential for change within states over time. For example, cognitive impairment during infancy might not be defined equivalently to cognitive impairment during childhood, adolescence or adulthood. Each separate development stage could require separate development states.

Economic evaluations using DAMs to model child development might not be possible given the information required to populate the mode. Parameters would be required regarding the probability of children transitioning from and into each of the individual development states across all life stages and health related utilities, healthcare costs, and cross-sectoral costs and benefits would need to be assigned to each development state. Because of these complexities it might be more appropriate to estimate lifetime effects of early life health technologies using direct or indirect sources of empirical evidence.

7.5 Research Recommendations and Application of Methodology

The following recommendations are intended to inform researchers when estimating lifetime effects of early life circumstances for economic evaluation. The recommendations indicate how methodologies adopted/advanced in this thesis might be applied in other research scenarios:

1. A decision perspective should be specified together with the relevant decision endpoints for this perspective. The economic evaluation of UK child health technologies/interventions might follow methods in this thesis and adopt two perspectives conducting analyses from: (i) a health centric decision maker's perspective using QALYs and healthcare costs as decision endpoints; and (ii) a cross-sectoral perspective including QALYs, healthcare costs and monetary returns to other sectors as decision endpoints. Ultimately which perspective is presented as primary or secondary should depend on the decision maker who the evaluation is intended to inform.
2. The early life circumstances and relevant comparator should be specified *prior* to the conduct of primary research. This will be self-evident for trial based economic evaluations which require evidence regarding the incremental lifetime effects of each treatment vs. the relevant control. The relevant early life circumstance may be less obvious when conducting economic evaluation using a DAM. It might not be appropriate/possible to model the development process as states within the DAM. It might be more appropriate to attach empirical estimates regarding lifetime effects to health states already defined within the model. For example, the DAM in this research contained health states for children exposed

to postnatal depression, and children not exposed to postnatal depression, thus requiring evidence regarding the incremental lifetime effects for children exposed to postnatal depression symptoms (compared with children not exposed). Each different DAM will have specific disease states that determine the type of evidence and comparison required.

3. Lifetime effects could be estimated by directly observing the relationship between early life circumstance and final decision endpoint. Direct evidence is unlikely to be available from RCTs, may be available from cmRCT designs (trials within cohorts), but is most likely to be available from birth cohorts who have reached adulthood. For the UK, relevant longitudinal cohorts include the 1970 British Cohort Study, the 1958 National Child Development study, and the 1946 National Survey of Health & Development.
4. Lifetime effects could also be estimated through indirect estimation by linking data across multiple studies. The first stage of indirect estimation requires evidence about the effect of early life circumstances on intermediate measures of child development. If possible, primary research should identify effects on development measures in time series data. Primary data could be obtained by conducting relatively long-term experimental studies which utilise data from promising new sources such as the Born in Bradford's Better Start experimental birth cohort. Equally, as was demonstrated in this thesis, observational evidence can be obtained from recent birth cohorts such as the Millennium Cohort and the (original) Born in Bradford studies. As described in chapter three, growth curve models provide an appropriate method when establishing longitudinal effects on child development measures. Secondary evidence sources may be available and should be searched to determine whether an evidence synthesis through meta-analysis is possible. Ideally, the effect sizes identified in primary research would be pooled with all the available secondary data in a meta-analysis.
5. The second stage of indirect estimation requires the effects on intermediate outcomes to be translated to effects on lifetime outcomes. The mathematical model in chapter five might be useful for other analysts when translating effects. This would require the effects of early life circumstances to be measured on equivalent intermediate endpoints as used as model input parameters in chapter five, i.e. General Cognitive Ability at age ten, and the Rutter Behaviour Scale, or the Strength and Difficulties Questionnaire at age ten. New mathematical models are likely to be developed in the future so researchers should search the contemporary literature base. It would be most appropriate to obtain predictions for each of the available models.
6. The final economic evaluation should establish the cost-effective strategy alongside the associated probability of cost-effectiveness. Decision makers are likely to benefit from

results with and without lifetime parameters to illustrate how their inclusion affects the cost-effectiveness decision. Similarly, results should be presented for all perspectives being investigated. In addition, appropriate figures should be included that illustrate cost-effectiveness results across a range of values for the cost-effectiveness threshold (k) and the consumption value for health (v). If predictions have been obtained from different models, they might be presented as different scenario analyses, or could be combined into a single estimate using model averaging. Scenario analyses/model averaging could also be used to address any other structural assumption made during indirect estimation, for example the assumption of increasing, decreasing or consistent effect sizes over time.

7.6 Conclusions

This thesis addresses a limitation common to economic evaluations of childhood health technologies. A technology may affect child development and subsequently influence lifetime health and economic outcomes, but these lifetime effects may be excluded from an analysis in the absence of typical sources of direct evidence. The research in this thesis describes and advances methods for indirect estimation by linking data from two empirical studies and applies these methods to estimate lifetime effects in children exposed to symptoms of postnatal depression. The lifetime effects estimates were used to update an existing UK based economic evaluation assessing the cost-effectiveness of screening for postnatal depression.

The influence of lifetime effects on policy recommendations was demonstrated as their inclusion affected the cost-effective screening strategy.

The findings from this thesis suggest that if economic evaluations exclude lifetime effects, they risk providing a partial analysis and may underestimate the benefits of childhood health technologies if the technology fosters lifespan development.

The methods advanced in this research could be applied when assessing the cost-effectiveness of other child health technologies. Methods for indirect estimation could reduce the risk of bias in decision making which may occur if lifetime effects are routinely excluded from economic evaluation.

Appendices

Chapter 3 Appendices

Appendix 3.1 Maternal Depression Questionnaire Items

Appendix 3.1A: The 9-Item Rutter Malaise Index

Questionnaire Item	Responses (and scoring)
Do your feel tired most of the time?	Yes (=1)/ No (=0)
Do you often feel miserable or depressed?	Yes (=1)/ No (=0)
Do you often get worried about things?	Yes (=1)/ No (=0)
Do you often get into violent rage?	Yes (=1)/ No (=0)
Do you often suddenly become scared for no good reason?	Yes (=1)/ No (=0)
Are you easily upset or irritated?	Yes (=1)/ No (=0)
Are you constantly keyed up and jittery?	Yes (=1)/ No (=0)
Does every little thing get on your nerves and wear you out?	Yes (=1)/ No (=0)
Does your heart often race like mad?	Yes (=1)/ No (=0)

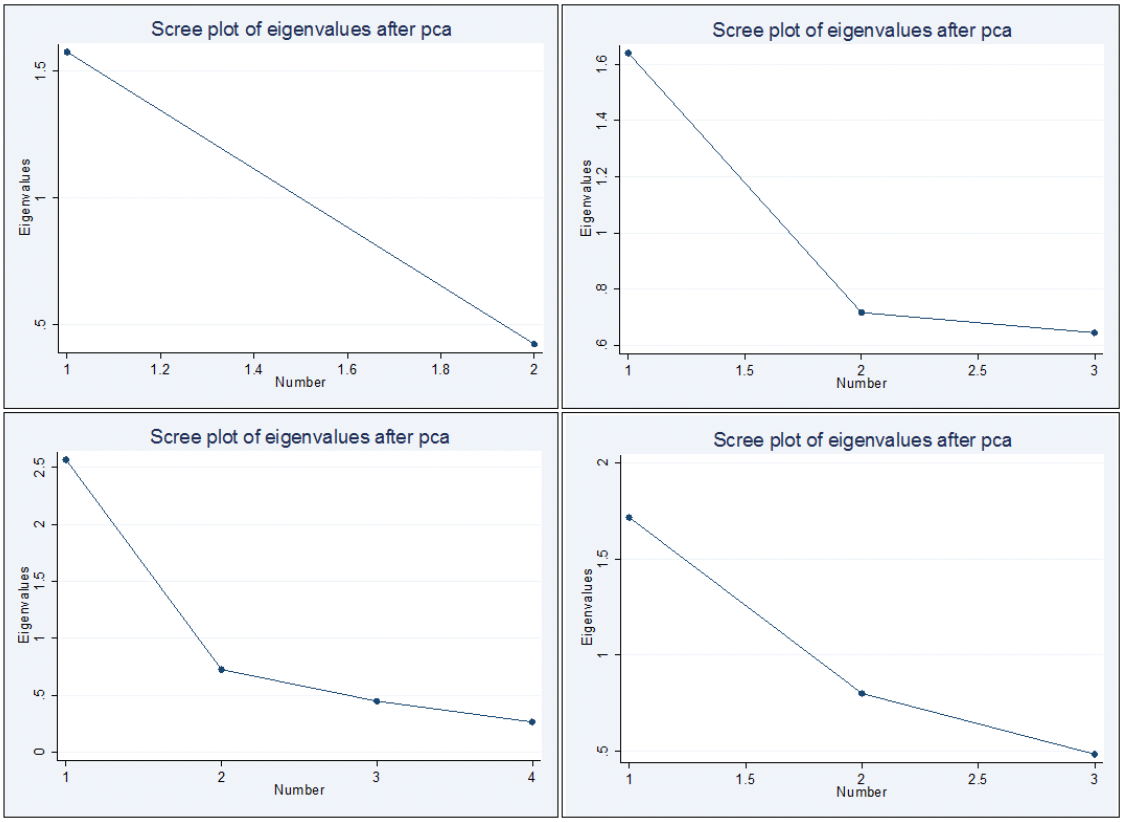
Notes: Administered during sweep one of the Millennium Cohort Study.
Total score are derive by summing responses for each item, scores >=4 indicate symptoms of postnatal depression,

Appendix 3.1B: The Kessler 6 Index

Questionnaire Item	Responses (and scoring)
During the last 30 days, how often did you feel so depressed that nothing could cheer you up?	None of the time (=0); a little of the time (=1); some of the time (=2); most of the time (=3); all of the time (=4).
During the last 30 days, about how often did you feel hopeless?	None of the time (=0); a little of the time (=1); some of the time (=2); most of the time (=3); all of the time (=4).
During the last 30 days, about how often did you feel restless or fidgety?	None of the time (=0); a little of the time (=1); some of the time (=2); most of the time (=3); all of the time (=4).
During the last 30 days, about how often did you feel that everything was an effort?	None of the time (=0); a little of the time (=1); some of the time (=2); most of the time (=3); all of the time (=4).
During the last 30 days, about how often did you feel worthless?	None of the time (=0); a little of the time (=1); some of the time (=2); most of the time (=3); all of the time (=4).
During the last 30 days, about how often did you feel nervous?	None of the time (=0); a little of the time (=1); some of the time (=2); most of the time (=3); all of the time (=4).

Notes: Administered during sweeps 2-5 of the Millennium Cohort Study.
Total scores are derived by summing responses for each item, scores >=13 indicate symptoms of maternal psychological distress.

Appendix 3.2: Scree Plots from Principal Component Analysis



Notes: Scree Plots of eignevalues for unrotated factors identified in PCA for sweep two (top left), sweep three (top right), sweep four (bottom left) and sweep five (bottom right).

Appendix 3.3: Full Results Growth Curve Model 1.

Appendix 3.3A: SQRT SDQ Total Difficulties Model

	Coefficient	Standard Error	P-Value
Maternal Depression Variables			
Postnatal Depression Symptoms	0.386	0.033	0.000
Maternal Distress Symptoms	0.253	0.021	0.000
Covariates			
Male Child	0.242	0.021	0.000
Child Ethnicity (ref. White British)			
Mixed Race	0.021	0.071	0.766
Indian	-0.236	0.088	0.007
Pakistani/Bangladeshi	-0.051	0.075	0.496
Black	-0.039	0.085	0.644
Other	0.006	0.148	0.966
Siblings (ref. =0)			
1	-0.020	0.016	0.198
2	-0.353	0.021	0.101
3+	-0.006	0.030	0.837
Maternal Age at Child Birth	-0.019	0.002	0.000
Low Household Income	0.132	0.031	0.000
Maternal Education (ref. NVQ 2)			
NVQ Level 1	0.026	0.047	0.581
NVQ Level 3	-0.075	0.032	0.019
NVQ Level 4	-0.188	0.027	0.000
NVQ Level 5	-0.249	0.037	0.000
Overseas Qualification	0.187	0.081	0.021
No Qualifications	0.180	0.049	0.000
Single Mother	0.139	0.023	0.000
Mother in Employment	-0.041	0.010	0.000
Time Metrics			
Age at SDQ Assessment (months)	0.010	0.001	0.000
Age^2	0.001	0.000	0.000
Time Interaction Terms (with Age)			
Postnatal Depression Symptoms	0.000	0.000	0.644
Maternal Distress Symptoms	-0.002	0.001	0.054
Male Child	-0.001	0.001	0.138
Child Ethnicity (ref. White British)			
Mixed Race	-0.002	0.002	0.134
Indian	-0.004	0.002	0.089
Pakistani/Bangladeshi	-0.001	0.002	0.714
Black	-0.003	0.002	0.191
Other	-0.007	0.004	0.080
Maternal Age at Child Birth	0.000	0.000	0.076
Low Household Income	0.005	0.001	0.000

Appendix 3.3A: SQRT SDQ Total Difficulties Model *continued*

	Coefficient	Standard Error	P-Value
Time Interaction Terms (with Age) *continued*			
Maternal Education (ref. NVQ 2)			
NVQ Level 1	-0.003	0.001	0.047
NVQ Level 3	-0.000	0.001	0.800
NVQ Level 4	-0.001	0.001	0.072
NVQ Level 5	-0.005	0.001	0.000
Overseas Qualification	-0.003	0.002	0.158
No Qualifications	-0.001	0.001	0.259
Single Mother	-0.001	0.001	0.317
Time Interaction Terms (with Age^2)			
Postnatal Depression Symptoms	0.0003	0.0008	0.722
Maternal Distress Symptoms	-0.0005	0.0010	0.616
Male Child	-0.0017	0.0006	0.003
Child Ethnicity (ref. White British)			
Mixed Race	-0.0002	0.0001	0.196
Indian	-0.0000	0.0002	0.827
Pakistani/Bangladeshi	0.0003	0.0002	0.120
Black	-0.0002	0.0003	0.348
Other	-0.0009	0.0001	0.096
Maternal Age at Child Birth	0.0009	0.0006	0.123
Low Household Income	0.0004	0.0000	0.000
Maternal Education (ref. NVQ 2)			
NVQ Level 1	-0.0002	0.0001	0.188
NVQ Level 3	-0.0004	0.0009	0.655
NVQ Level 4	-0.0001	0.0000	0.020
NVQ Level 5	-0.0006	0.0001	0.000
Overseas Qualification	-0.0005	0.0002	0.058
No Qualifications	-0.0000	0.0001	0.447
Single Mother	-0.0002	0.0000	0.016

Appendix 3.3B: GCA Model

	Coefficient	Standard Error	P-Value
Maternal Depression Variables			
Postnatal Depression Symptoms	-0.274	0.843	0.673
Maternal Distress Symptoms	-1.832	0.116	0.000
Covariates			
Male Child	-0.588	0.2950	0.049
Child Ethnicity (ref. White British)			
Mixed Race	0.484	1.082	0.655
Indian	2.333	1.324	0.078
Pakistani/Bangladeshi	-3.278	1.141	0.004
Black	-4.076	1.283	0.001
Other	2.965	2.224	0.182
Siblings (ref. =0)			
1	-1.502	0.366	0.000
2	-3.224	0.507	0.000
3+	-4.363	0.737	0.000
Maternal Age at Child Birth	0.336	0.033	0.000
Low Household Income	-1.216	0.545	0.026
Maternal Education (ref. NVQ 2)			
NVQ Level 1	-1.911	0.725	0.008
NVQ Level 3	1.012	0.496	0.041
NVQ Level 4	3.847	0.411	0.000
NVQ Level 5	5.151	0.575	0.000
Overseas Qualification	-2.435	1.303	0.062
No Qualifications	-4.369	0.781	0.000
Single Mother	-0.718	0.211	0.001
Mother in Employment	0.670	0.465	0.328
Time Metrics (Dummy Variables)			
Sweep 2	105.948	0.339	0.000
Sweep 3	104.056	0.348	0.000
Sweep 4	101.865	0.374	0.000
Sweep 5	101.136	0.439	0.000
Time Interaction Terms (with Sweep 2)			
Postnatal Depression Symptoms	-0.381	0.556	0.543
Maternal Distress Symptoms	-0.218	0.746	0.448
Male Child	-3.200	0.359	0.000
Child Ethnicity (ref. White British)			
Mixed Race	-0.601	1.216	0.621
Indian	-8.285	1.488	0.000
Pakistani/Bangladeshi	-9.398	1.275	0.000
Black	-2.223	1.438	0.122
Other	-9.656	2.526	0.000
Siblings (ref. =0)			
1	-2.219	0.418	0.000
2	-3.706	0.580	0.000
3+	-4.679	0.840	0.000
Maternal Age at Child Birth	-0.027	0.037	0.468
Low Household Income	-0.853	0.615	0.165

Appendix 3.3B: GCA Model *continued*

	Coefficient	Standard Error	P-Value
Time Interaction Terms (with Sweep 2) *continued*			
Maternal Education (ref. NVQ 2)			
NVQ Level 1	-0.249	0.813	0.759
NVQ Level 3	0.525	0.576	0.363
NVQ Level 4	-0.136	0.481	0.777
NVQ Level 5	-0.187	0.797	0.814
Overseas Qualification	-0.618	1.502	0.681
No Qualifications	0.586	0.857	0.494
Time Interaction Terms (with Sweep 3)			
Postnatal Depression Symptoms	-0.009	0.525	0.601
Maternal Distress Symptoms	-0.078	0.500	0.853
Male Child	-1.405	0.336	0.000
Child Ethnicity (ref. White British)			
Mixed Race	-1.231	1.115	0.284
Indian	-3.831	1.405	0.006
Pakistani/Bangladeshi	-4.271	1.206	0.000
Black	-1.107	1.363	0.417
Other	-5.595	2.373	0.018
Siblings (ref. =0)			
1	-0.644	0.392	0.100
2	-0.544	0.544	0.317
3+	-2.306	0.791	0.004
Maternal Age at Child Birth	-0.056	0.035	0.112
Low Household Income	0.847	0.602	0.159
Maternal Education (ref. NVQ 2)			
NVQ Level 1	-0.372	0.777	0.632
NVQ Level 3	-0.502	0.546	0.357
NVQ Level 4	-1.220	0.454	0.007
NVQ Level 5	-1.564	0.703	0.026
Overseas Qualification	0.233	1.437	0.871
No Qualifications	0.390	0.824	0.636
Time Interaction Terms (with Sweep 4)			
Postnatal Depression Symptoms	-0.762	0.410	0.045
Maternal Distress Symptoms	0.242	0.101	0.002
Male Child	-0.348	0.326	0.285
Child Ethnicity (ref. White British)			
Mixed Race	-0.537	1.116	0.631
Indian	2.116	1.363	0.120
Pakistani/Bangladeshi	3.700	1.182	0.002
Black	1.636	1.333	0.219
Other	1.285	2.314	0.579
Siblings (ref. =0)			
1	0.205	0.380	0.588
2	-0.434	0.527	0.411
3+	-1.042	0.769	0.176

Appendix 3.3B: GCA Model *continued*

	Coefficient	Standard Error	P-Value
Time Interaction Terms (with Sweep 4) *continued*			
Maternal Age at Child Birth	-0.027	0.034	0.066
Low Household Income	-0.737	0.603	0.222
Maternal Education (ref. NVQ 2)			
NVQ Level 1	-1.718	0.763	0.024
NVQ Level 3	0.250	0.531	0.638
NVQ Level 4	-0.440	0.440	0.318
NVQ Level 5	-0.516	0.657	0.432
Overseas Qualification	0.381	1.409	0.787
No Qualifications	0.175	0.0.814	0.830

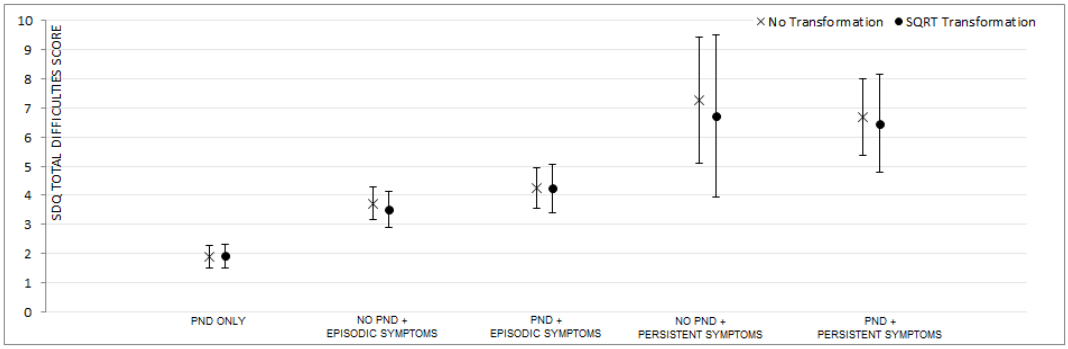
Appendix 3.4: Results of Growth Curve Model 2.

Appendix 3.4A: Sweep 5 Mean Incremental Effects and 95% CI for Growth Curve Model 2

Maternal Depression Trajectory	SQRT SDQ Total Difficulties*		SDQ Total Difficulties		GCA	
	Mean	*95% CI*	*Mean*	*95% CI*	*Mean*	*95% CI*
PND Symptoms + No Later Symptoms	1.90	(1.50, 2.31)	1.88	(1.48, 2.28)	-0.16	(-1.62, 1.31)
No PND Symptoms + Later Episodic Symptoms	3.51	(2.89, 4.14)	3.71	(3.16, 4.27)	-4.94	(-6.93, -2.96)
PND Symptoms + Later Episodic Symptoms	4.23	(3.41, 5.05)	4.26	(3.56, 4.96)	-2.39	(-4.93, 0.14)
No PND + Later Persistent Symptoms	6.72	(3.94, 9.51)	7.25	(5.08, 9.42)	-4.22	(-11.40, 2.97)
PND + Later Persistent Symptoms	6.47	(4.77, 8.16)	6.68	(5.37, 7.99)	-5.90	(-10.59, -1.21)

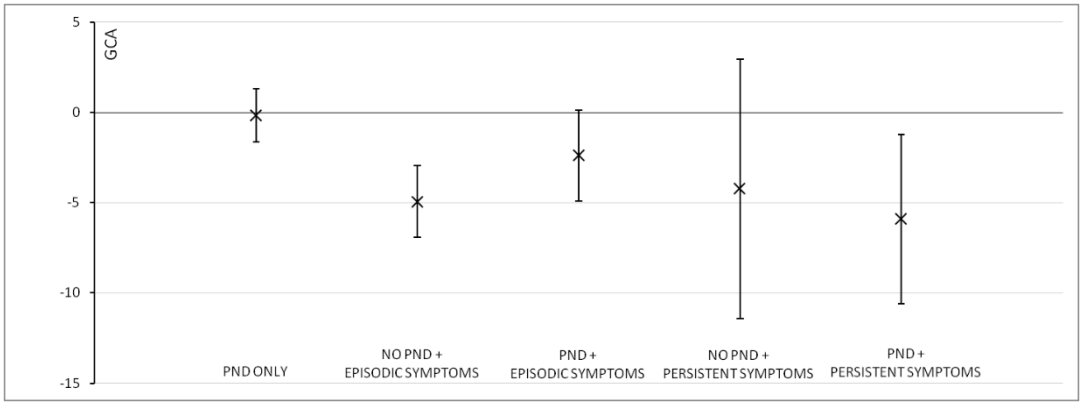
Notes: Incremental effects in children exposed to symptoms by maternal depression trajectory.
PND=postnatal depression symptoms identified during sweep one.
Later episodic depression= 1 or 2 episodes of maternal depression symptoms during sweeps 2-5.
Later persistent depression= 3 or 4 episodes of maternal depression symptoms during sweeps 2-5.
All incremental effects vs. trajectory with: no PND symptoms + no later symptoms.
*Results back transformed to raw scale.
SDQ n= 7,9701; GCA n= 4,270

Appendix 3.4B: Sweep 5 SDQ Total Difficulties Mean Incremental Effects and 95% CI



Notes: Figure plots incremental effect sizes and 95% confidence intervals for children aged 11 (sweep five) whose mothers follow different trajectories for distress symptoms. All incremental effects are vs. no symptoms of distress for any sweep. SDQ Total Difficulties outcomes are reported on the raw scale for the non-transformed outcome (crosses) and on the back transformed scale for the square root outcome (dots).

Appendix 3.4C: Sweep 5 GCA Mean Incremental Effects and 95% CI



Notes: Figure plots incremental effect sizes and 95% confidence intervals on GCA outcomes for children aged 11 (sweep five) whose mothers follow different trajectories for distress symptoms. All incremental effects are vs. no symptoms of distress for any sweep.

Appendix 3.5: Mean "Moderate" Threshold Results at Sweep 5

	SQRT SDQ Total Difficulties		GCA	
	Moderate[1]	*Severe*[2]	*Moderate*[1]	*Severe*[2]
Growth Curve Model 1 Results				
Cumulative Distress	0.21***	0.25***	-0.80***	-1.83***
Growth Curve Model 2 Results				
PND Only	0.26**	0.37***	-1.03	-0.16
PND & Episodic Symptoms	0.51***	0.65***	-2.65	-4.94***
No PND & Persistent Symptoms	0.59***	0.76***	-0.74***	-2.39
PND & Persistent Symptoms	0.85***	1.14***	-3.40***	-4.22
No PND & Persistent Symptoms	0.99***	1.10***	-3.46***	-5.90***

Notes: All results are reported for outcomes at sweep five (age eleven) as β coefficients.
Compares results of Growth Curve Model 1 & Growth Curve Model 2 using different threshold values for the Kessler 6 Instrument.
1: Moderate symptoms are defined as scores >=5
2: Severe symptoms are scores >=13.
***p<0.001, **p<0.01, *p<0.05.

Chapter 4 Appendices

Appendix 4.1: Search Strategies

Appendix 4.1A: EMBASE Databases Search Strategy

N	Search Term
1	((develop$ or grow$ or matur$ or inance$ or inanc$ or evolu$) adj2 (human$ or child$ or juvenile$ or youth$ or youngster$ or adolescen$ or kindergart$ or pre school$ or preschool$ or newborn$ or new born$ or neonat$ or toddler$ or teen$ or kid or kids or infant$ or paediatric$ or pediatric$ or fetus$ or foetus$ or fetal$ or foetal$)).ti,ab.
2	exp child development/ or exp fetus development/ or exp human development/ or exp postnatal development/ or exp adolescent development/ or exp child growth/
3	(exp "growth, development and aging"/ or exp language development/ or exp "growth, development and aging disorders"/ or exp psychomotor development/ or exp physical development/ or exp nervous system development/ or exp cognitive development/ or exp speech development/ or exp motor development/ or exp mental development/ or exp brain maturation/) and (child$ or juvenile$ or youth$ or youngster$ or adolescen$ or kindergart$ or pre school$ or preschool$ or newborn$ or new born$ or neonat$ or toddler$ or teen$ or kid or kids or infant$ or paediatric$ or pediatric$ or fetus$ or foetus$ or fetal$ or foetal$).ti,ab.
4	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or theor$ or simulat$ or concept$ or decision$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or prognostic$ or prognosis$ or crystal ball or crystalball or cost$ or econom et$ or economic$ or monte carlo or montecarlo or informat$ or inance$ or disease$ or physical or inancey$ or process$ or experiment$ or informat$ or explan$ or explain$) adj5 model$).ti,ab.
5	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or theor$ or simulat$ or concept$ or decision$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or prognostic$ or prognosis$ or crystal ball or crystalball or cost$ or econom et$ or economic$ or monte carlo or montecarlo or informat$ or disease$ or informat$ or explan$ or explain$ or life cycle or lifecycle) adj5 framework$).ti,ab.
6	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or simulat$ or concept$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or monte carlo or montecarlo or explain$ or explan$ or lifecycle or life cycle) adj2 (outcome or outcomes or input or inputs or output or outputs or variab$ or function$)).ti,ab.
7	exp individual based population model/ or exp information model/ or exp statistical model/ or exp physical model/ or exp stochastic model/ or exp experimental model/ or exp computer model/ or exp mathematical model/ or exp biological model/ or exp psychological model/ or exp nonbiological model/ or exp compartment model/ or exp educational model/ or exp disease model/ or exp process model/ or exp theoretical model/ or exp population model/ or exp genetic model/ or exp loglinear model/ or exp hidden Markov model/ or exp model/

Appendix 4.1A: *Continued*

N	Search Term
8	Socioeconomics
9	Cost benefit analysis/
10	Cost effectiveness analysis/
11	Cost of illness/
12	Cost control/
13	Economic aspect/
14	Financial management/
15	Health care cost/
16	Health care financing/
17	Health economics/
18	Hospital cost/
19	(fiscal or financial or finance or funding).tw.
20	Cost minimization analysis/
21	(cost adj estimate$).mp.
22	(cost adj variable$).mp.
23	(unit adj cost$).mp.
24	exp economic aspect/ or costs.tw
25	(Economic or economics).tw.
26	1 or 2 or 3
27	4 or 5 or 6 or 7
28	Or/8-25
29	26 and 27 and 28
30	Limit 29 to (inance language and human and yr="2000 –Current")
31	30 not (human epidermal growth or human growth hormone or HGH).ti,ab.

Notes: NHS EED filter (Glanville et al., 2008) was applied to identify economic studies in EMBASE

Appendix 4.1B: Medline Database Search Strategy

N	Search Term
1	((human$ or child$ or juvenile$ or youth$ or youngster$ or adolescen$ or kindergart$ or pre school$ or preschool$ or newborn$ or new born$ or neonat$ or toddler$ or teen$ or kid or kids or infant$ or paediatric$ or pediatric$ or fetus$ or foetus$ or fetal$ or foetal$) adj2 (develop$ or grow$ or matur$ or inance$ or inanc$ or evolu$)).ti,ab.
2	exp Child Development/or exp Adolescent Development/ or exp Human Development/ or exp Embryonic Development/ or exp Fetal Development/
3	(exp Program Development/ or exp Developmental Disabilities/ or exp Language Development/ or exp Language Development Disorders/ or exp "Growth and Development"/ or exp Personality Development/) and (child$ or juvenile$ or youth$ or youngster$ or adolescen$ or kindergart$ or pre school$ or preschool$ or newborn$ or new born$ or neonat$ or toddler$ or teen$ or kid or kids or infant$ or paediatric$ or pediatric$ or fetus$ or foetus$ or fetal$ or foetal$).ti,ab.
4	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or theor$ or simulat$ or concept$ or decision$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or prognostic$ or prognosis$ or crystal ball or crystalball or cost$ or econom et$ or economic$ or monte carlo or montecarlo or informat$ or inance$ or disease$ or physical or inancey$ or process$ or experiment$ or informat$ or explan$ or explain$) adj5 model$).ti,ab.
5	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or theor$ or simulat$ or concept$ or decision$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or prognostic$ or prognosis$ or crystal ball or crystalball or cost$ or econom et$ or economic$ or monte carlo or montecarlo or informat$ or disease$ or informat$ or explan$ or explain$ or life cycle or lifecycle) adj5 framework$).ti,ab.
6	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or simulat$ or concept$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or monte carlo or montecarlo or explain$ or explan$ or lifecycle or life cycle) adj2 (outcome or outcomes or input or inputs or output or outputs or variab$ or function$)).ti,ab.
7	exp Models, Theoretical/ or exp Models, Biological/ or exp Models, Statistical/ or exp Computer Simulation/ or exp Linear Models/ or exp Models, Genetic/ or exp Models, Psychological/ or exp Markov Chains/ or exp Models, Economic/ or exp Models, Econometric/

Appendix 4.1B *continued*

N	Search Term
8	Economics/
9	"costs and cost analysis"/
10	Cost allocation/
11	Cost-benefit analysis/
12	Cost control/
13	Cost savings/
14	Cost of illness/
15	Cost sharing/
16	"deductibles and coinsurance"/
17	Medical savings accounts/
18	Health care costs/
19	Direct service costs/
20	Drug costs/
21	Employer health costs/
22	Hospital costs/
23	Health expenditures/
24	Capital expenditures/
25	Value of life/
26	Exp economics, hospital/
27	Exp economics, medical/
28	Economics, nursing/
29	Economics, pharmaceutical/
30	Exp "fees and charges"/
31	Exp budgets/
32	(low adj cost).mp.
32	(low adj cost).mp.
33	(high adj cost).mp.
34	(health?care adj cost$).mp.
35	(fiscal or funding or financial or finance).tw.
36	(cost adj estimate$).mp.
37	(cost adj variable).mp.
38	(unit adj cost$).mp.
39	(economic$ or pharmacoeconomic$ or price$ or pricing).tw.
40	or/8-39
41	exp economics/
42	cost:.mp.
43	1 or 2 or 3
44	Or/4-7
45	40 or 41 or 42
46	43 and 44 and 45
47	Limit 46 to (inance language and human and yr="2000 –Current")
48	47 not (human epidermal growth or human growth hormone or HGH).ti,ab.

Notes: The SIGN filter for economic studies (in Medline) was used as part of the search strategy (Glanville et al., 2008).

Appendix 4.1C: PSYCINFO Database Search Strategy

N	Search Term
1	((human$ or child$ or juvenile$ or youth$ or youngster$ or adolescen$ or kindergart$ or pre school$ or preschool$ or newborn$ or new born$ or neonat$ or toddler$ or teen$ or kid or kids or infant$ or paediatric$ or pediatric$ or fetus$ or foetus$ or fetal$ or foetal$) adj2 (develop$ or grow$ or matur$ or inance$ or inanc$ or evolu$)).ti,ab.
2	exp Neonatal Development/ or exp Human Development / or exp Adolescent Development/ or exp Early Childhood Development/ or exp Childhood Development/ or exp Prenatal Development/ or exp Infant Development/
3	(exp Physical Development/ or exp Intellectual Development Disorder/ or exp Perceptual Motor Development/ or exp Emotional Development/ or exp Language Development/ or exp Perceptual Development/ or exp Speech Development/ or exp Neural Development/ or exp Motor Development/ or exp Psychosexual Development/ or exp Psychosocial Development/ or exp Development/ or exp Reading Development/ or exp Ego Development/ or exp Intellectual Development/ or exp Cognitive Development/ or exp Psychomotor Development/ or exp Brain Development/ or exp Moral Development/ or exp Psychological Development/ or exp Personality Development/) *and* (child$ or juvenile$ or youth$ or youngster$ or adolescen$ or kindergart$ or pre school$ or preschool$ or newborn$ or new born$ or neonat$ or toddler$ or teen$ or kid or kids or infant$ or paediatric$ or pediatric$ or fetus$ or foetus$ or fetal$ or foetal$).ti,ab.
4	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or theor$ or simulat$ or concept$ or decision$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or prognostic$ or prognosis$ or crystal ball or crystalball or cost$ or econom$ or economic$ or monte carlo or montecarlo or informat$ or inance$ or disease$ or physical or inancey$ or process$ or experiment$ or informat$ or explan$ or explain$) adj5 model$).ti,ab.
5	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or theor$ or simulat$ or concept$ or decision$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or prognostic$ or prognosis$ or crystal ball or crystalball or cost$ or econom$ or economic$ or monte carlo or montecarlo or informat$ or disease$ or informat$ or explan$ or explain$ or life cycle or lifecycle) adj5 framework$).ti,ab.
6	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or simulat$ or concept$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or monte carlo or montecarlo or explain$ or explan$ or lifecycle or life cycle) adj2 (outcome or outcomes or input or inputs or output or outputs or variab$ or function$)).ti,ab.
7	exp Clinical Models/ or exp Models/ or exp Decision Making/ or exp Mathematical Modeling/ or exp Computer Software/ or exp Simulation/ or exp Statistical Analysis/ or exp Prediction/ or exp Theories/ or exp Theoretical Interpretation/ or exp Statistical Estimation/ or exp Multivariate Analysis/ or exp Psychological Theories/

Appendix 4.1C: *continued*

N	Search Term
8	"costs and cost analysis"/
9	"Cost Containment"/
10	(economic adj2 evaluation$).ti,ab.
11	(economic adj2 analy$).ti,ab.
12	(economic adj2 (study or studies)).ti,ab.
13	(cost adj2 evaluation$).ti,ab.
14	(cost adj2 analy$).ti,ab.
15	(cost adj2 (study or studies)).ti,ab.
16	(cost adj2 effective$).ti,ab.
17	(cost adj2 benefit$).ti,ab.
18	(cost adj2 utili$).ti,ab.
19	(cost adj2 minimi$).ti,ab.
20	(cost adj2 consequence$).ti,ab.
21	(cost adj2 comparison$).ti,ab.
22	(cost adj2 identificat$).ti,ab.
23	(pharmacoeconomic$ or pharmaco-economic$).ti,ab.
24	exp "Costs and Cost Analysis"/ or exp "Cost Containment"/ or exp "Response Cost"/ or exp Finance/ or exp Salaries/ or exp Human Capital/ or exp Health Education/ or exp Resource Allocation/ or exp "Income (Economic)"/ or exp Health Care Services/ or exp Health Care Costs/ or exp Economy/ or exp Health Care Utilization/ or exp Economics/ or exp Financial Strain/ or exp Health Care Delivery/ or exp Financial Services/ or exp Government/ or exp Academic Achievement/ or exp Government Policy Making/ or exp Budgets/ or exp Behavioral Economics/ or exp Health Care Economics/ or exp Evolutionary Economics/
25	(Economic$ or Cost$ or Econometric or (fiscal or financial or finance or funding) or Socioeconomic$).ti,ab.
26	or/8-25
27	(task adj2 cost$).ti,ab,id.
28	(switch$ adj2 cost$).ti,ab,id.
29	(metabolic adj cost).ti,ab,id.
30	((energy or oxygen) adj cost).ti,ab,id.
31	((energy or oxygen) adj expenditure).ti,ab,id.
32	Or/27-31
33	(animal or animals or rat or rats or mouse or mice or hamster or hamsters or dog or dogs or cat or cats or bovine or sheep or ovine or pig or pigs).ab,ti,id,de.
34	editorial.dt.
35	letter.dt.
36	dissertation abstract.pt.
37	or/33-36
38	(0003-4819 or 0003-9926 or 0959-8146 or 0098-7484 or 0140-6736 or 0028-4793 or 1469-493X).is.
39	26 not (32 or 37 or 38)
40	1 or 2 or 3
41	4 or 5 or 6 or 7
42	39 and 40 and 41
43	limit 42 to (human and inance language and yr="2000 –Current")

Notes: The NHS CRD NHS EED search filter was used as the economics search filter. As this search filter is designed only to obtain economic evaluations additional search terms were used as part of the "economics" search term. Key search terms were also identified from SIGN filters (Glanville et al., 2008) and from McKinlay et al. (2006) (both designed for EMBASE/MEDLINE and translated to the PSYCINFO database).

Appendix 4.1D: ECONLIT Database Search Strategy

N	Search Term
1	((human$ or child$ or juvenile$ or youth$ or youngster$ or adolescen$ or kindergart$ or pre school$ or preschool$ or newborn$ or new born$ or neonat$ or toddler$ or teen$ or kid or kids or infant$ or paediatric$ or pediatric$ or fetus$ or foetus$ or fetal$ or foetal$) adj2 (develop$ or grow$ or matur$ or inance$ or inanc$ or evolu$)).ti,ab.
2	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or theor$ or simulat$ or concept$ or decision$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or prognostic$ or prognosis$ or crystal ball or crystalball or cost$ or econom$ or economic$ or monte carlo or montecarlo or informat$ or inance$ or disease$ or physical or inancey$ or process$ or experiment$ or informat$ or explan$ or explain$) adj5 model$).ti,ab..
3	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or theor$ or simulat$ or concept$ or decision$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or prognostic$ or prognosis$ or crystal ball or crystalball or cost$ or econom$ or economic$ or monte carlo or montecarlo or informat$ or disease$ or informat$ or explan$ or explain$ or life cycle or lifecycle) adj5 framework$).ti,ab.
4	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or simulat$ or concept$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or monte carlo or montecarlo or explain$ or explan$ or lifecycle or life cycle) adj2 (outcome or outcomes or input or inputs or output or outputs or variab$ or function$)).ti,ab.
5	2 or 3 or 4
6	1 and 5
7	limit 6 to (yr="2000 –Current" and inance)
8	limit 7 to dissertations
9	7 not 8

Notes: As ECONLIT is a database for economic articles the economics filter was removed from the search strategy. MeSH headings and exploded terms were not used as these were not supported in this database.

Appendix 4.1E: Maternity and Infant Care Database Search Strategy

N	Search Term
1	(develop$ or grow$ or matur$ or inance$ or inanc$ or evolu$).ti,ab.
2	(Development or Growth or Outcome – long term or Child development or Neurodevelopmental outcome or "Infant – growth and development"or Human Development or Fetal development or Infant development or Infant growth or Child growth or "Adolescent – growth and development" or Fetal growth).de.
3	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or theor$ or simulat$ or concept$ or decision$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or prognostic$ or prognosis$ or crystal ball or crystalball or cost$ or econom$ or economic$ or monte carlo or montecarlo or informat$ or inance$ or disease$ or physical or inancey$ or process$ or experiment$ or informat$ or explan$ or explain$) adj5 model$).ti,ab.
4	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or theor$ or simulat$ or concept$ or decision$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or prognostic$ or prognosis$ or crystal ball or crystalball or cost$ or econom$ or economic$ or monte carlo or montecarlo or informat$ or disease$ or informat$ or explan$ or explain$ or life cycle or lifecycle) adj5 framework$).ti,ab.
5	((mathematic$ or maths or math or numeric$ or number$ or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili$ or stochastic or empiric$ or simulat$ or concept$ or statistic$ or comput$ or analytic$ or data or inanc$ or markov$ or inance$ or predict$ or forecast$ or monte carlo or montecarlo or explain$ or explan$ or lifecycle or life cycle) adj2 (outcome or outcomes or input or inputs or output or outputs or variab$ or function$)).ti,ab.
6	(Models or Long term outcome or Models – statistical).de.
7	(Cost-benefit analysis or Economics or "Costs and cost analysis" or Insurance or Insurance – liability or Finance or Health care costs – statistics or "Cost savings" or Socioeconomic factors).de.
8	(Economic$ or econom$ or economy or cost$ or deductible$ or coinsurance or insur$ or savings or expenditure$ or socioeconomic$ or inance$ or fiscal or funding or pricing or price$ or salary or salaries or capital or allocation or income or incomes or budget$ or money or monetary or profit$).ti,ab.
9	1 or 2
9	3 or 4 or 5 or 6
10	7 or 8
11	9 and 10 and 11
12	limit 12 to yr="2000 –Current"

Notes: As the maternity and infant care database is specifically for publications relating to children and mothers the "child" component of the "child" "development" term was omitted.

Appendix 4.1F: Child Development & Adolescent Studies Database (Ebsco Host) Search Strategy

N	Search Term
1	TI (develop* or grow* or matur* or inance* or inanc* or evolu*) or AB (develop* or grow* or matur* or inance* or inanc* or evolu*)
2	TI (mathematic* or maths or math or numeric* or number* or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili* or stochastic or empiric* or theor* or simulat* or concept* or decision* or statistic* or comput* or analytic* or data or inanc* or markov* or inance* or predict* or forecast* or prognostic* or prognosis* or crystal ball or crystalball or cost* or economet* or economic* or monte carlo or montecarlo or informat* or inance* or disease* or physical or inancey* or process* or experiment* or informat* or explan* or explain* or lifecycle or life cycle) N5 TI (Model*)
3	AB (mathematic* or maths or math or numeric* or number* or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili* or stochastic or empiric* or theor* or simulat* or concept* or decision* or statistic* or comput* or analytic* or data or inanc* or markov* or inance* or predict* or forecast* or prognostic* or prognosis* or crystal ball or crystalball or cost* or economet* or economic* or monte carlo or montecarlo or informat* or inance* or disease* or physical or inancey* or process* or experiment* or informat* or explan* or explain* or lifecycle or life cycle) N5 AB (Model*)
4	TI (mathematic* or maths or math or numeric* or number* or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili* or stochastic or empiric* or theor* or simulat* or concept* or decision* or statistic* or comput* or analytic* or data or inanc* or markov* or inance* or predict* or forecast* or prognostic* or prognosis* or crystal ball or crystalball or cost* or economet* or economic* or monte carlo or montecarlo or informat* or disease* or informat* or explan* or explain* or life cycle or lifecycle) N5 TI (framework*)
5	AB (mathematic* or maths or math or numeric* or number* or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili* or stochastic or empiric* or theor* or simulat* or concept* or decision* or statistic* or comput* or analytic* or data or inanc* or markov* or inance* or predict* or forecast* or prognostic* or prognosis* or crystal ball or crystalball or cost* or economet* or economic* or monte carlo or montecarlo or informat* or disease* or informat* or explan* or explain* or life cycle or lifecycle) N5 AB (framework*)
6	TI (mathematic* or maths or math or numeric* or number* or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili* or stochastic or empiric* or simulat* or concept* or statistic* or comput* or analytic* or data or inanc* or markov* or inance* or predict* or forecast* or monte carlo or montecarlo or explain* or explan* or lifecycle or life cycle) N2 TI (outcome or outcomes or input or inputs or output or outputs or variab* or function*)
7	AB (mathematic* or maths or math or numeric* or number* or linear or nonlinear or non linear or loglinear or log linear or deterministic or probabili* or stochastic or empiric* or simulat* or concept* or statistic* or comput* or analytic* or data or inanc* or markov* or inance* or predict* or forecast* or monte carlo or montecarlo or explain* or explan* or lifecycle or life cycle) N2 AB (outcome or outcomes or input or inputs or output or outputs or variab* or function*)
8	TI (Economic* or economet* or economy or cost* or deductible* or coinsurance or insur* or savings or expenditure* or socioeconomic* or inance* or fiscal or funding or pricing or price* or salary or salaries or capital or allocation or income or incomes or budget* or money or monetary or profit*)

Appendix 4.1F: *continued*

N	Search Term
9	AB (Economic* or economet* or economy or cost* or deductible* or coinsurance or insur* or savings or expenditure* or socioeconomic* or inance* or fiscal or funding or pricing or price* or salary or salaries or capital or allocation or income or incomes or budget* or money or monetary or profit*)
10	2 or 3 or 4 or 5 or 6 or 7
11	8 or 9
12	1 and 10 and 11

Notes: As a child related database the "child" component was omitted from the "child" "development" term. Despite the database being specifically for child development, the "development" term was required as results were not specific enough with this term omitted. It was not possible to include additional limitations (e.g. English language only) as these were not supported in the database.

Appendix 4.2: Screening Forms

Appendix 4.2A: Screening Form 1: Title & Abstracts

Has the search strategy identified the correct development term (e.g. economic development, development of the model)?
Yes (include)
No (exclude)
Can't tell (include)

Does the study report an empirical model?
Yes (include)
No (exclude)
Can't tell (include)

Is the study sample for a diseased population?
Yes (exclude)
No (include)
Can't tell (exclude)

Are any model outcomes reported at age 16 or above?
Yes (include)
No (exclude)
Can't tell (include)

Are there any model inputs at age 12 or below?
Yes (include)
No (exclude)
Can't tell (include)

Is the outcome relevant to health economic evaluation?
Yes (include)
No (exclude)
Can't Tell (include)

Is the model outcome specific to a single disease/health condition?
Yes (exclude)
No (include)
Can't tell (include)

Is the study an animal study?
Yes (exclude)
No (include)
Can't tell (include)

Is the study written in a language other than English?
Yes (include)
No (exclude)
Can't tell (include)

Is the study published after the 1st January 2000?
Yes (include)
No (exclude)
Can't tell (include)

Notes: Studies were excluded if any of the below criteria resulted in an "exclude" answer.

Appendix 4.2B: Screening Form 2: Full Texts

(For objective 1 only):
Does the model include QALYs and Healthcare Costs as outcome variables?
Yes (include)
No (exclude)

(For objectives 1 & 2):
Is there any discussion of the child development literature?
Yes (include)
No (exclude)

Does the study report an empirical model?
Yes (include)
No (exclude)

Are any model outcomes reported at age 16 or above?
Yes (include)
No (exclude)

Are there any model inputs at age 12 or below?
Yes (include)
No (exclude)

Is the outcome relevant to health economic evaluation?
Yes (include)
No (exclude)

Is the study sample a diseased population only?
Yes (exclude)
No (include)

Is the model outcome specific to a single disease/health condition?
Yes (exclude)
No (include)

Notes: Studies were excluded if any of the below criteria resulted in an "exclude" answer.

Appendix 4.3: Data Charting Form

Primary Author	Year	Location	Discipline	Individual Development Domains	Environmental Parameters	Model Type	Outcome Sectors	Time Horizon (years)
Adair, L.	2013	Multiple	Health Sciences	Physical	Family economic	GLM regression	Education & health	31
Boyle, M.	2007	Canada	Human Development	Cognitive, socioemotional & physical	Family economic, family psychosocial & neighbourhood	Hierarchical linear model	Education	30
Chandola, T.	2006	UK	Epidemiology	Cognitive & physical	Family economic	Structural equation model	Health	42
Chartier, M.	2010	Canada	Health Sciences	None	Family economic & family psychosocial	GLM regression	Education & health	Not stated
Creel, M.	2006	USA	Economics	Cognitive	Family economic	Quadratic parametric model	Education	11
Cunha, F.	2008	USA	Economics	Cognitive & socioemotional	Family economic & family psychosocial	Dynamic factor model	Education & employment	22
Cunha, F.	2010	USA	Economics	Cognitive & socioemotional	Family economic & family psychosocial	Dynamic factor model	Crime & education	Not stated
Daniels, M.	2004	Philippines	Other	Cognitive & physical	Family economic & family psychosocial	GLM regression	Education	17
Dishion, T.	2010	USA	Psychology	Cognitive & socioemotional	None	Structural equation model	Crime	8
Dubow, E.	2009	USA	Psychology	Cognitive & socioemotional	Family economic & family psychosocial	Structural equation model	Education & employment	11
Engle, P.	2011	Multiple	Human Development	Socioemotional	School/peer & neighbourhood	GLM regression	Employment	Not stated

Appendix 4.3 *continued*

Primary Author	Year	Location	Discipline	Individual Development Domains	Environmental Parameters	Model Type	Outcome Sectors	Time Horizon (years)
Fergusson, D.	2004	New Zealand	Psychology	Socioemotional	Family economic, family psychosocial & school/peer	GLM regression	Crime	11
Ferrer, E.	2004	USA	Psychology	Cognitive	None	Growth curve model	Education	30
Friedman, H.	2014	USA	Psychology	Socioemotional	None	Structural equation model	Education & health	44
Frijters, P.	2010	UK	Economics	Physical	Family economic & neighbourhood	Cox proportional hazard model	Health	Not stated
Hagger-Johnson, G.	2012	UK	Psychology	Cognitive	Family economic	Structural equation model	Health	60
Hagger-Johnson, G.	2011	UK	Epidemiology	Cognitive	Family economic, family psychosocial	Structural equation model	Employment & health	42
Hampson, S.	2015	USA	Psychology	Socioemotional	None	Structural equation Model	Health	41
Hatch, S.	2010	UK	Psychology	Cognitive, socioemotional & physical	Family economic & family psychosocial	GLM regression	Education & health	33
Healey, A.	2004	UK	Economics	Cognitive & socioemotional	Family economic & family psychosocial	Dynamic factor analysis	Crime, education & employment	24
Herrenkohl, T.	2010	USA	Sociology	Socioemotional	Family economic & family psychosocial	Structural equation model	Crime & health	20
Hertzman, C.	2001	UK	Epidemiology	Cognitive, socioemotional & physical	Family economic & family psychosocial	GLM regression	Health	31
Huang, C.	2011	Mexico	Public Health	Physical	Family economic & neighbourhood	GLM regression	Education, employment & health	40
Jimerson, S.	2001	USA	Psychology	Cognitive & socioemotional	Family economic, family psychosocial & school/peer	Hierarchical linear model	Education	14

Appendix 4.3 *continued*

Primary Author	Year	Location	Discipline	Individual Development Domains	Environmental Parameters	Model Type	Outcome Sectors	Time Horizon (years)
Johnson, W.	2006	USA	Human Development	Cognitive, socioemotional & physical	None	Hierarchical linear model	Education	6
Kuh, D.	2004	UK	Epidemiology	Cognitive	Family economic & family psychosocial	Cox proportional hazard model	Health	45
Kuh, D.	2009	UK	Epidemiology	Cognitive, socioemotional, physical	Family economic & family psychosocial	Cox proportional hazard model	Health	54
Lacourse E.	2002	Canada	Human Development	Socioemotional	Family economic	GLM regression	Crime	12
Lacourse, E.	2006	Canada	Health Sciences	None	None	Hierarchical linear model	Crime	Not stated
Lager, A.	2012	Sweden	Health Sciences	Cognitive	Family economic	Structural equation model	Health	68
Layard, R.	2014	UK	Economics	Cognitive, socioemotional, physical	Family economic & family psychosocial	GLM regression	Education, employment & health	29
Lee, T.	2013	USA	Human Development	None	Family economic & family psychosocial	Structural equation model	Health	Not stated
Lindeboom, M.	2006	UK	Economics	Cognitive, socioemotional & physical	Family economic	GLM regression	Employment & health	42
Martens, P.	2014	Canada	Public Health	None	Family economic, family psychosocial & neighbourhood	GLM regression	Education	Not stated
Mason, P.	2010	USA	Psychology	Socioemotional	Family economic	Growth curve model	Crime	14
Mason, W.	2007	USA	Economics	None	Family economic	GLM regression	Employment	Not stated
Moffit, T.	2011	New Zealand	Psychology	Cognitive & socioemotional	Family economic	GLM regression	Crime, employment & health	29
Moody-Ayers, S.	2007	USA	Epidemiology	None	Family economic	GLM regression	Health	Not stated
Muennig, P.	2009	USA	Economics	Cognitive	Family economic & family psychosocial	Unsure	Crime, employment & health	Not stated

Appendix 4.3 *continued*

Primary Author	Year	Location	Discipline	Individual development domains	Environmental Parameters	Model type	Outcome Sectors	Time Horizon (years)
Nandi, A.	2014	USA	Epidemiology	Physical	Family economic	Structural equation model	Health	Not stated
Nikulina, V.	2011	USA	Psychology	None	Family economic, family psychosocial & neighbourhood	Hierarchical linear model	Crime, education & health	Not stated
Petras, H.	2005	USA	Public Health	Cognitive & socioemotional	Family economic, school/peer & neighbourhood	General growth mixture model	Crime & health	15
Reynolds, A.	2011	USA	Human Development	Cognitive & socioemotional	Family economic & family psychosocial	Structural equation model	Crime, employment & health	21
Risi, S.	2003	USA	Psychology	Cognitive & socioemotional	School/peer	GLM regression	Education	13
Roeser, R.	2003	USA	Psychology	Cognitive, socioemotional & physical	Family economic	Structural equation model	Education	8
Rosa Dias, P.	2009	UK	Economics	Physical	Family economic	Unsure	Health	46
Savage, J.	2002	Multiple	Psychology	None	Family psychosocial & neighbourhood	GLM regression	Crime	Not stated
Schoon, I.	2003	UK	Psychology	Socioemotional & physical	Family economic & family psychosocial	Structural equation model	Health	28
Sharieff, W.	2008	Pakistan	Economics	Cognitive & physical	None	Unsure	Employment	55
Shen, K	2014	China	Epidemiology	Physical	Family economic & neighbourhood	Structural equation model	Education, employment & health	65
Slopen, N.	2014	USA	Psychology	Cognitive, socioemotional & physical	Family economic, family psychosocial & neighbourhood	GLM regression	Health	Not stated

Appendix 4.3 *continued*

Primary Author	Year	Location	Discipline	Individual Development Domains	Environmental Parameters	Model Type	Outcome Sectors	Time Horizon (years)
Te Velde, S.	2008	Netherlands	Economics	Socioemotional	None	Unsure	Health	Not stated
Tran, B.	2015	Canada	Public Health	Physical	None	Growth curve model	Health	60
Tubeuf, S.	2013	UK	Health Sciences	Physical	Family economic & family psychosocial	GLM regression	Education & health	46
Vartanian, T.	2005	USA	Sociology	None	Family economic & neighbourhood	GLM regression	Employment	Not stated
Wodtke, G.	2016	USA	Sociology	None	Family economic, family psychosocial & neighbourhood	Two stage regression	Education	Not stated
World Bank	2003	Multiple	Economics	Socioemotional	Family economic, family psychosocial & school/peer	GLM regression	Crime, education, employment & health	Not stated

Chapter 5 Appendices

Appendix 5.1: Items on the Rutter 'A' Behaviour Scale

Questionnaire Item:	Responses (and scoring)
Very restless. Often running about or jumping up and down. Hardly ever still.	No (=0), somewhat (=1), certainly applies (=2)
Is squirmy or fidgety	No (=0), somewhat (=1), certainly applies (=2)
Often destroys own or others' belongings	No (=0), somewhat (=1), certainly applies (=2)
Frequently fights other children	No (=0), somewhat (=1), certainly applies (=2)
Not much liked by other children	No (=0), somewhat (=1), certainly applies (=2)
Often worried, worries about many things	No (=0), somewhat (=1), certainly applies (=2)
Tends to do things on his/her own – rather solitary	No (=0), somewhat (=1), certainly applies (=2)
Irritable. Is quick to fly off the handle	No (=0), somewhat (=1), certainly applies (=2)
Often appears miserable, unhappy, tearful or distressed	No (=0), somewhat (=1), certainly applies (=2)
Sometimes takes things belonging to others	No (=0), somewhat (=1), certainly applies (=2)
Has twitches, mannerisms or tics of the face or body	No (=0), somewhat (=1), certainly applies (=2)
Frequently sucks thumb or finger	No (=0), somewhat (=1), certainly applies (=2)
Frequently bites nails or fingers	No (=0), somewhat (=1), certainly applies (=2)
Is often disobedient	No (=0), somewhat (=1), certainly applies (=2)
Cannot settle to anything for more than a few moments	No (=0), somewhat (=1), certainly applies (=2)
Tends to be fearful or afraid of new things or new situations	No (=0), somewhat (=1), certainly applies (=2)
Is over fussy or over particular	No (=0), somewhat (=1), certainly applies (=2)
Often tells lies	No (=0), somewhat (=1), certainly applies (=2)
Bullies other children	No (=0), somewhat (=1), certainly applies (=2)

Notes: The 19-Item Rutter 'A' Behaviour Scale. The questionnaire was administered to participants' mothers for the age 10 sweep. Each mother was presented with a statement (questionnaire item) and asked how often this statement applied to their child (response). Total scores were obtained by summing all response items.

Appendix 5.2: Description of Statistical Tools used to Inform Model Selection

The *modified Park's test* assess the relationship between the variance and mean of a selected GLM distribution (Jones, 2010). As described by Jones (2010), the relationship between variance and mean in GLM regression "implies:

$$\ln[\, \mathrm{var}(\, y_i \mid x_i \,)\,] = \ln(\alpha) + \upsilon \ln(\mu_i)$$

The [Park] test exploits this by regressing $\ln[(y_i - \hat{y_i})^2]$ on $\ln(\hat{y_i})$ and a constant, typically using a GLM to estimate the model, having tested for the appropriate form of the link function to use". The modified Park test provides guidance on the appropriate distribution according to the slope coefficient for $\ln(\hat{y_i})$ from the regression equation (Jones, 2010). Slopes equal to: 0 suggest a Gaussain, 1 a Poisson, 2 a gamma, and 3 an inverse Guassian distribution.

Pregibon's link test is often used to identify a range of possible link functions given the associated GLM distribution. The Pregibon test assesses the suitability of a link by (i) conducting a GLM with the hypothesized link function; (ii) generating the predicted value, $\hat{y}$, and the predicted value squared, $\hat{y}^2$, of the outcome variables from the GLM; and (iii) estimating a second GLM regression that is identical to the first but including $\hat{y}$ and $\hat{y}^2$ as explanatory variables. The Pregibon test identifies a link function as incorrectly specified if the $\hat{y}^2$ coefficient is statistically significant ($p<0.05$) (Jones, 2010), (Pregibon, 1980).

The *Akaike Information Criterion* (AIC) is a statistic that estimates the amount of information that is lost from a given model based on the number of model parameters and the models maximum likelihood function. The AIC establishes model desirability according to goodness of fit and level of overfitting, where a lower AIC value is indicative of a more desirable model. As suggested by Jones and O'Donnell (2002), AIC values can be used to select the most appropriate model when comparing a range of different candidate models that are no necessarily nested. These including GLM, transformed OLS and two part regression models.

The *Vuong test* compares the fit of two non-nested models estimated in same dataset using a maximum likelihood function by testing null hypotheses that the models fit the data equally well (Desmarais and Harden, 2013). The Vuong test is commonly applied in to test whether a single part Poission or two part zero inflation model better fits count data. Desmarais and Harden (2013) suggest the default Vuong test in STATA can produce biased results as it does not correct for the increased number of parameters in two part models. The authors provide a bias corrected STATA command which was used in this research.

Appendix 5.3: Marginal Effects and 95% CI for Transformed Rutter Behavioural Scale

	Within Study (3.5%)	Lifetime (3.5%)	Lifetime (1.5%)	Lifetime (0 %)
QALYs				
GCA	0.001	0.003	0.007	0.019
	(0.001, 0.002)	(0.001, 0.004)	(0.004, 0.010)	(0.010, 0.027)
RBS	-0.002	-0.004	-0.011	-0.031
	(-0.004, -0.000)	(-0.008, -0.001)	(-0.020, -0.002)	(-0.057, -0.005)
Healthcare Costs				
GCA	-£8.06	-£15.81	-£40.63	-£133.58
	(-£14.50, -£1.63)	(-£28.96, -£2.66)	(-£75.40, -£5.85)	(-£249.67, -£17.49)
RBS	£28.78	£58.76	£154.71	£515.33
	(£7.13, £50.43)	(£14.61, £102.91)	(£32.57, £276.85)	(£127.30, £903.37)
Returns to Education*				
GCA	£2,514.17	£4,524.81	£10,895.39	£21,781.50
	(£1,611.30, £3,419.10)	(£3,153.16, £5,899.61)	(£7,839.96, £13,958.04)	(£15,913.75, £24,690.36)
RBS	-£638.38	-£625.13	-£873.82	-£1,068.46
	(-£3,283.21, £2,024.72	(-£4,641.40, £3,418.54)	(-£9,817.45, £8,131.26)	(-£18,240.27, £16,222.67)
Crime Costs				
GCA	-£0.92	-£1.00	-£1.29	-£2.02
	(-£1.55, -£0.30)	(-£1.68, -£0.32)	(-£2.16, -£0.42)	(-£3.38, -£0.65)
RBS	£2.98	£3.24	£4.16	£6.52
	(£0.99, £4.98	(£1.07, £5.41)	(£1.38, £6.95)	(£2.16, £10.89)

Notes: Marginal effects of General Cognitive Ability and the Rutter Behavioural Scale with measurement scale transformed to range from 0-40.
Each marginal effect is per unit increase in GCA and (transformed) RBS.
Outcomes obtained according to time horizon (study/lifetime) and discount rate.
*Results were back transformed to the raw scale using Duan's smearing estimator.

Appendix 5.4: Predicted Mean Lifetime Effects by Childhood Exposure to Maternal Depression Trajectory

Maternal Depression Trajectory	Childhood		Adulthood			
	GCA	SDQ/ Rutter	QALYs	Healthcare Costs	Returns to Education	Crime Costs
PND Symptoms + No Later Symptoms	-0.16	1.90	-0.008	£114.17	-£1,911.72	£6.32
No PND Symptoms + Later Episodic Symptoms	-4.94	3.51	-0.029	£284.35	-£24,546.77	£16.31
PND Symptoms + Later Episodic Symptoms	-2.39	4.23	-0.024	£286.34	-£13,458.60	£16.10
No PND Symptoms + Later Persistent Symptoms	-4.22	6.72	-0.040	£461.59	-£23,295.57	£25.99
PND Symptoms + Later Persistent Symptoms	-5.90	6.47	-0.044	£473.46	-£30,740.97	£26.86

Notes: Predicted mean lifetime outcomes (discounted at 3.5%) for children grouped according to maternal depression trajectory.
All are incremental costs/benefits compared with a "no depression symptoms" control group.
GCA and SDQ scores were taken from chapter three, Growth Curve Model 2.

Chapter 6 Appendices

Appendix 6.1: Deterministic Cost-effectiveness Results: Health Centric Decision Perspective & Maternal Outcomes

Strategy	QALYs	Healthcare Costs	ICERs vs. Lower Costing Strategy
EPDS (cutoff 7)	0.84486	£188.17	Dominated
EPDS (cutoff 8)	0.84604	£175.76	Dominated
EPDS (cutoff 9)	0.84709	£160.40	Dominated
EPDS (cutoff 10)	0.84764	£151.95	Dominated
PHQ9 (cutoff 9/10)	0.84774	£145.90	Dominated
EPDS (cutoff 11)	0.84815	£139.47	Dominated
EPDS (cutoff 12)	0.84835	£134.26	Dominated
EPDS (cutoff 13)	0.84825	£133.94	Dominated
EPDS (cutoff 14)	0.84838	£122.85	Dominated
Whooley + EPDS (cutoff 7)	0.84811	£117.49	Dominated
Whooley + EPDS (cutoff 8)	0.84852	£113.15	Dominated
EPDS (cutoff 15)	0.84833	£112.71	Dominated
EPDS (cutoff 16)	0.84831	£107.05	Dominated
Whooley + EPDS (cutoff 9)	0.84880	£106.08	Dominated
Whooley + EPDS (cutoff 10)	0.84893	£102.10	Dominated
Whooely + PHQ9 (cutoff 9/10)	0.84886	£98.06	Dominated
Whooley + EPDS (cutoff 11)	0.84896	£95.01	Dominated
Whooley + EPDS (cutoff 12)	0.84897	£92.00	£37,621
Whooley + EPDS (cutoff 13)	0.84890	£91.35	Extendedly Dominated
Whooley + EPDS (cutoff 14)	0.84875	£83.78	£32,137
Standard Care	0.84606	£82.07	Dominated
Whooley + EPDS (cutoff 15)	0.84851	£76.16	Extendedly Dominated
Whooley + EPDS (cutoff 16)	0.84838	£71.94	

Appendix 6.2: Deterministic Cost-effectiveness Results: Health Centric Decision Perspective & Maternal and Child Outcomes

Strategy	QALYs	Healthcare Costs	ICERs vs. Lower Costing Strategy
EPDS (cutoff 7)	0.8442	£193.18	Dominated
EPDS (cutoff 8)	0.8454	£180.76	Dominated
EPDS (cutoff 9)	0.8465	£165.54	Dominated
EPDS (cutoff 10)	0.8470	£157.17	Dominated
PHQ-9 (cutoff 9/10)	0.8471	£151.26	Dominated
EPDS (cutoff 11)	0.8475	£144.88	Dominated
EPDS (cutoff 12)	0.8477	£139.77	Dominated
EPDS (cutoff 13)	0.8476	£139.49	Dominated
EPDS (cutoff 14)	0.8477	£128.68	Dominated
Whooley + EPDS (cutoff 7)	0.8475	£122.60	Dominated
EPDS (cutoff 15)	0.8476	£118.85	Dominated
Whooley + EPDS (cutoff 8)	0.8479	£118.26	Dominated
EPDS (cutoff 16)	0.8475	£113.36	Dominated
Whooley + EPDS (cutoff 9)	0.8482	£111.31	Dominated
Whooley + EPDS (cutoff 10)	0.8483	£107.40	Dominated
Whooely + PHQ9 (cutoff 9/10)	0.8482	£103.50	Dominated
Whooley + EPDS (cutoff 11)	0.8483	£100.51	£520,564.14
Whooley + EPDS (cutoff 12)	0.8483	£97.58	£30,914.30
Whooley + EPDS (cutoff 13)	0.8482	£96.97	Extendedly Dominated
Whooley + EPDS (cutoff 14)	0.8480	£89.67	£26,817.93
Standard Care	0.8453	£87.95	Dominated
Whooley + EPDS (cutoff 15)	0.8478	£82.34	Extendedly Dominated
Whooley + EPDS (cutoff 16)	0.8476	£78.28	

Notes: QALYs and healthcare costs discounted at 3.5%

Appendix 6.3: Deterministic Cost-effectiveness Results: Cross-Sectoral Decision Perspective & Maternal and Child Outcomes

Strategy	QALYs	Healthcare Costs	Education Costs	NCB k=£20,000 v=£20,000	NCB k=£30,000 v=£30,000
EPDS (cutoff 7)	0.8442	£193.18	£100.11	-109.59	-120.54
EPDS (cutoff 8)	0.8454	£180.76	£100.10	-73.41	-72.48
EPDS (cutoff 9)	0.8465	£165.54	£102.68	-40.24	-29.04
EPDS (cutoff 10)	0.8470	£157.17	£104.24	-22.60	-6.00
EPDS (cutoff 11)	0.8475	£144.88	£108.39	-4.72	16.76
EPDS (cutoff 12)	0.8477	£139.77	£110.20	2.35	25.71
EPDS (cutoff 13)	0.8476	£139.49	£111.02	-0.36	21.92
EPDS (cutoff 14)	0.8477	£128.68	£116.63	6.89	30.19
EPDS (cutoff 15)	0.8476	£118.85	£122.84	8.77	31.20
EPDS (cutoff 16)	0.8475	£113.36	£126.25	10.01	32.03
PHQ9 (cuttof 9/10)	0.8471	£151.26	£107.17	-17.58	0.05
Whooley + EPDS (cutoff 7)	0.8475	£122.60	£102.10	23.78	45.22
Whooley + EPDS (cutoff 8)	0.8479	£118.26	£102.09	36.45	62.06
Whooley + EPDS (cutoff 9)	0.8482	£111.31	£104.54	46.09	74.27
Whooley + EPDS (cutoff 10)	0.8483	£107.40	£106.03	51.07	80.53
Whooley + EPDS (cutoff 11)	0.8483	£100.51	£109.96	54.18	83.70
Whooley + EPDS (cutoff 12)	0.8483	£97.58	£111.68	**55.27**	**84.74**
Whooley + EPDS (cutoff 13)	0.8482	£96.97	£112.46	53.70	82.46
Whooley + EPDS (cutoff 14)	0.8480	£89.67	£117.80	51.95	78.86
Whooley + EPDS (cutoff 15)	0.8478	£82.34	£123.69	47.88	72.03
Whooley + EPDS (cutoff 16)	0.8476	£78.28	£126.93	45.70	68.37
Whooely + PHQ9 (cutoff 9/10)	0.8482	£103.50	£108.80	50.45	79.04
Standard Care	0.8453	£87.95	£117.64	0.00	0.00

Notes: NCB is net consumption benefit. The cost-effective strategy is marked in **bold**.

Appendix 6.4: Deterministic Sensitivity Analysis Cost-effectiveness Results

Appendix 6.4A Utility Decrement for False Positive Mother's (from 2% to 0%)

Strategy	QALYs	Costs	ICERs vs. lower costing strategy
EPDS (cutoff 7)	0.850041	£188.17	Dominated
EPDS (cutoff 8)	0.850042	£175.76	£527,537.76
EPDS (cutoff 9)	0.849888	£160.40	Dominated
EPDS (cutoff 10)	0.849795	£151.95	Dominated
PHQ9 (cutoff 9/10)	0.849620	£145.90	Dominated
EPDS (cutoff 11)	0.849548	£139.47	Dominated
EPDS (cutoff 12)	0.849440	£134.26	Dominated
EPDS (cutoff 13)	0.849391	£133.94	Dominated
EPDS (cutoff 14)	0.849056	£122.85	Dominated
Whooley + EPDS (cutoff 7)	0.849923	£117.49	Dominated
Whooley + EPDS (cutoff 8)	0.849923	£113.15	£48,303.18
EPDS (cutoff 15)	0.848686	£112.71	Dominated
EPDS (cutoff 16)	0.848482	£107.05	Dominated
Whooley + EPDS (cutoff 9)	0.849777	£106.08	£45,032.05
Whooley + EPDS (cutoff 10)	0.849688	£102.10	£30,212.54
Whooely + PHQ9 (cutoff 9/10)	0.849523	£98.06	Extendedly Dominated
Whooley + EPDS (cutoff 11)	0.849454	£95.01	£29,322.47
Whooley + EPDS (cutoff 12)	0.849351	£92.00	£27,944.62
Whooley + EPDS (cutoff 13)	0.849305	£91.35	Extendedly Dominated
Whooley + EPDS (cutoff 14)	0.848987	£83.78	Dominated
Standard Care	0.848996	£82.07	£18,287.60
Whooley + EPDS (cutoff 15)	0.848635	£76.16	Extendedly Dominated
Whooley + EPDS (cutoff 16)	0.848442	£71.94	

Notes: Results for health centric perspective with maternal outcomes only.

Appendix 6.4B Prevalence of Postnatal Depression =20%

Strategy	QALYs	Costs	ICERs vs. lower costing strategy
EPDS (cutoff 7)	0.8326	£274.12	Dominated
EPDS (cutoff 8)	0.8336	£263.26	Dominated
EPDS (cutoff 9)	0.8343	£245.78	Dominated
EPDS (cutoff 10)	0.8346	£235.95	Dominated
PHQ9 (cutoff 9/10)	0.8344	£226.09	Dominated
EPDS (cutoff 11)	0.8347	£218.56	Dominated
EPDS (cutoff 12)	0.8348	£211.18	Dominated
EPDS (cutoff 13)	0.8346	£209.63	Dominated
Whooley + EPDS (cutoff 7)	0.8352	£209.22	Dominated
Whooley + EPDS (cutoff 8)	0.8356	£205.43	Dominated
Whooley + EPDS (cutoff 9)	0.8356	£195.42	£998,449.63
EPDS (cutoff 14)	0.8343	£191.16	Dominated
Whooley + EPDS (cutoff 10)	0.8356	£189.62	£40,259.60
Whooely + PHQ9 (cutoff 9/10)	0.8353	£181.76	Extendedly Dominated
Whooley + EPDS (cutoff 11)	0.8353	£177.28	£37,522.09
EPDS (cutoff 15)	0.8337	£172.62	Dominated
Whooley + EPDS (cutoff 12)	0.8352	£171.96	£23,543.98
Whooley + EPDS (cutoff 13)	0.8351	£170.19	Extendedly Dominated
EPDS (cutoff 16)	0.8334	£162.35	Dominated
Whooley + EPDS (cutoff 14)	0.8345	£155.25	£22,403.24
Standard Care	0.8321	£145.87	Dominated
Whooley + EPDS (cutoff 15)	0.8338	£139.40	Extendedly Dominated
Whooley + EPDS (cutoff 16)	0.8334	£130.64	

Notes: Results for health centric perspective with maternal outcomes only.

Appendix 6.4C Lifetime Effects Proportionate to Duration of Exposure to Depression Symptoms

Strategy	QALYs	Costs	ICERs vs. lower costing strategy
EPDS (cutoff 7)	0.8441	£194.34	Dominated
EPDS (cutoff 8)	0.8453	£181.92	Dominated
EPDS (cutoff 9)	0.8463	£166.64	Dominated
EPDS (cutoff 10)	0.8469	£158.23	Dominated
PHQ9 (cutoff 9/10)	0.8470	£152.25	Dominated
EPDS (cutoff 11)	0.8474	£145.85	Dominated
EPDS (cutoff 12)	0.8476	£140.69	Dominated
EPDS (cutoff 13)	0.8475	£140.39	Dominated
EPDS (cutoff 14)	0.8476	£129.44	Dominated
Whooley + EPDS (cutoff 7)	0.8473	£123.71	Dominated
EPDS (cutoff 15)	0.8475	£119.46	Dominated
Whooley + EPDS (cutoff 8)	0.8478	£119.37	Dominated
EPDS (cutoff 16)	0.8475	£113.88	Dominated
Whooley + EPDS (cutoff 9)	0.8480	£112.36	Dominated
Whooley + EPDS (cutoff 10)	0.84816	£108.42	Dominated
Whooely + PHQ9 (cutoff 9/10)	0.84808	£104.45	Dominated
Whooley + EPDS (cutoff 11)	0.84818	£101.43	£6,297,314.62
Whooley + EPDS (cutoff 12)	0.84818	£98.46	£33,917.61
Whooley + EPDS (cutoff 13)	0.84811	£97.84	Extendedly Dominated
Whooley + EPDS (cutoff 14)	0.84794	£90.40	£29,242.09
Standard Care	0.84525	£88.69	Dominated
Whooley + EPDS (cutoff 15)	0.84768	£82.93	£29,711.51
Whooley + EPDS (cutoff 16)	0.84754	£78.79	

Notes: Results for health centric perspective with maternal and child outcomes discounted at 3.5%.

Appendix 6.4D Effects of Postnatal Depression on Child Development Reduce by 25%

Strategy	QALYs	Costs	ICERs vs. lower costing strategy
EPDS (cutoff 7)	0.84440	£191.92	Dominated
EPDS (cutoff 8)	0.84558	£179.51	Dominated
EPDS (cutoff 9)	0.84661	£164.25	Dominated
EPDS (cutoff 10)	0.84716	£155.86	Dominated
PHQ9 (cutoff 9/10)	0.84724	£149.92	Dominated
EPDS (cutoff 11)	0.84765	£143.53	Dominated
EPDS (cutoff 12)	0.84784	£138.39	Dominated
EPDS (cutoff 13)	0.84774	£138.11	Dominated
EPDS (cutoff 14)	0.84785	£127.23	Dominated
Whooley + EPDS (cutoff 7)	0.84764	£121.32	Dominated
EPDS (cutoff 15)	0.84777	£117.32	Dominated
Whooley + EPDS (cutoff 8)	0.84805	£116.98	Dominated
EPDS (cutoff 16)	0.84773	£111.78	Dominated
Whooley + EPDS (cutoff 9)	0.84832	£110.00	Dominated
Whooley + EPDS (cutoff 10)	0.84845	£106.08	Dominated
Whooely + PHQ9 (cutoff 9/10)	0.84836	£102.14	Dominated
Whooley + EPDS (cutoff 11)	0.84846	£99.13	£989,596.05
Whooley + EPDS (cutoff 12)	0.84846	£96.19	£32,399.74
Whooley + EPDS (cutoff 13)	0.84839	£95.57	Extendedly Dominated
Whooley + EPDS (cutoff 14)	0.84821	£88.20	£28,030.44
Standard Care	0.84552	£86.48	Dominated
Whooley + EPDS (cutoff 15)	0.84794	£80.80	Extendedly Dominated
Whooley + EPDS (cutoff 16)	0.84780	£76.69	

Notes: Results for health centric perspective with maternal and child outcomes discounted at 3.5%.

Appendix 6.4E Effects of Postnatal Depression on Child Development Increase by 25%

Strategy	QALYs	Costs	ICERs vs. lower costing strategy
EPDS (cutoff 7)	0.84409	£194.43	Dominated
EPDS (cutoff 8)	0.84528	£182.01	Dominated
EPDS (cutoff 9)	0.84630	£166.82	Dominated
EPDS (cutoff 10)	0.84684	£158.47	Dominated
PHQ9 (cutoff 9/10)	0.84691	£152.60	Dominated
EPDS (cutoff 11)	0.84732	£146.24	Dominated
EPDS (cutoff 12)	0.84750	£141.15	Dominated
EPDS (cutoff 13)	0.84739	£140.88	Dominated
EPDS (cutoff 14)	0.84749	£130.14	Dominated
Whooley + EPDS (cutoff 7)	0.84732	£123.87	Dominated
EPDS (cutoff 15)	0.84739	£120.39	Dominated
Whooley + EPDS (cutoff 8)	0.84774	£119.53	Dominated
EPDS (cutoff 16)	0.84735	£114.94	Dominated
Whooley + EPDS (cutoff 9)	0.84799	£112.62	Dominated
Whooley + EPDS (cutoff 10)	0.84812	£108.73	Dominated
Whooely + PHQ9 (cutoff 9/10)	0.84803	£104.86	Dominated
Whooley + EPDS (cutoff 11)	0.84812	£101.88	£351,497.63
Whooley + EPDS (cutoff 12)	0.84811	£98.98	£29,533.85
Whooley + EPDS (cutoff 13)	0.84804	£98.38	Extendedly Dominated
Whooley + EPDS (cutoff 14)	0.84785	£91.14	£25,717.32
Standard Care	0.84516	£89.42	Dominated
Whooley + EPDS (cutoff 15)	0.84756	£83.89	Extendedly Dominated
Whooley + EPDS (cutoff 16)	0.84741	£79.87	

Notes: Results for health centric perspective with maternal and child outcomes discounted at 3.5%.

Appendix 6.4F Discount Rate Equal to 1.5% for Costs & QALYs

Strategy	QALYs	Healthcare Costs	Education Costs	Health Centric Perspective ICERs*	Cross-Sectoral Results: NCB k=£30,000 v=£30,000
EPDS (cutoff 7)	0.8434	£201.34	£189.11	Dominated	-99.09
EPDS (cutoff 8)	0.8446	£188.92	£189.09	Dominated	-51.02
EPDS (cutoff 9)	0.8456	£173.91	£193.97	Dominated	-10.74
EPDS (cutoff 10)	0.8461	£165.66	£196.91	Dominated	10.39
PHQ9 (cutoff 9/10)	0.8462	£159.99	£202.44	Dominated	12.22
EPDS (cutoff 11)	0.8466	£153.72	£204.74	Dominated	28.09
EPDS (cutoff 12)	0.8467	£148.76	£208.16	Dominated	34.82
EPDS (cutoff 13)	0.8466	£148.54	£209.70	Dominated	30.03
EPDS (cutoff 14)	0.8467	£138.19	£220.32	Dominated	31.42
Whooley + EPDS (cutoff 7)	0.8466	£130.92	£192.87	Dominated	64.24
EPDS (cutoff 15)	0.8465	£128.86	£232.03	Dominated	24.85
Whooley + EPDS (cutoff 8)	0.8470	£126.58	£192.85	Dominated	81.09
EPDS (cutoff 16)	0.8465	£123.81	£238.49	Dominated	21.49
Whooley + EPDS (cutoff 9)	0.8473	£119.83	£197.48	Dominated	90.30
Whooley + EPDS (cutoff 10)	0.8474	£116.05	£200.28	£251,855.94	**94.74**
Whooely + PHQ9 (cutoff 9/10)	0.8473	£112.37	£205.53	Dominated	89.85
Whooley + EPDS (cutoff 11)	0.8474	£109.47	£207.72	£138,200.67	93.10
Whooley + EPDS (cutoff 12)	0.8473	£106.69	£210.97	**£24,108.81**	92.03
Whooley + EPDS (cutoff 13)	0.8473	£106.14	£212.43	ED	88.81
Whooley + EPDS (cutoff 14)	0.8470	£99.27	£222.52	£21,064.12	78.67
Standard Care	0.8443	£97.54	£222.22	Dominated	0.00
Whooley + EPDS (cutoff 15)	0.8467	£92.43	£233.64	£20,554.74	64.63
Whooley + EPDS (cutoff 16)	0.8465	£88.69	£239.78		57.00

Notes: *ICERs calculated vs. next lower costing non-dominated or extendedly dominated strategy.
NCB is net consumption benefit. The cost-effective strategy at k=£30,000 is marked in **bold**.
ED= Extendedly Dominated

Appendix 6.4G Discount Rate Equal to 3.5% for Costs and 1.5% for QALYs

Strategy	QALYs	Healthcare Costs	Education Costs	Health Centric Perspective ICERs*	Cross-Sectoral Results: NCB k=£30,000 v=£30,000
EPDS (cutoff 7)	0.8434	£193.18	£100.11	Dominated	-116.10
EPDS (cutoff 8)	0.8446	£180.76	£100.10	Dominated	-68.04
EPDS (cutoff 9)	0.8456	£165.54	£102.68	Dominated	-25.26
EPDS (cutoff 10)	0.8461	£157.17	£104.24	Dominated	-2.61
PHQ-9 (cutoff 9/10)	0.8462	£151.26	£107.17	Dominated	2.05
EPDS (cutoff 11)	0.8466	£144.88	£108.39	Dominated	19.10
EPDS (cutoff 12)	0.8467	£139.77	£110.20	Dominated	27.59
EPDS (cutoff 13)	0.8466	£139.49	£111.02	Dominated	23.60
EPDS (cutoff 14)	0.8467	£128.68	£116.63	Dominated	30.45
Whooley + EPDS (cutoff 7)	0.8466	£122.60	£102.10	Dominated	49.16
EPDS (cutoff 15)	0.8465	£118.85	£122.84	Dominated	29.89
Whooley + EPDS (cutoff 8)	0.8470	£118.26	£102.09	Dominated	65.99
EPDS (cutoff 16)	0.8465	£113.36	£126.25	Dominated	29.85
Whooley + EPDS (cutoff 9)	0.8473	£111.31	£104.54	Dominated	77.59
Whooley + EPDS (cutoff 10)	0.8474	£107.40	£106.03	£264,148.60	83.47
Whooely + PHQ9 (cutoff 9/10)	0.8473	£103.50	£108.80	Dominated	81.27
Whooley + EPDS (cutoff 11)	0.8474	£100.51	£109.96	£145,167.41	85.65
Whooley + EPDS (cutoff 12)	0.8473	£97.58	£111.68	**£25,729.17**	**86.24**
Whooley + EPDS (cutoff 13)	0.8473	£96.97	£112.46	ED	83.78
Whooley + EPDS (cutoff 14)	0.8470	£89.67	£117.80	£22,541.81	78.82
Standard Care	0.8443	£87.95	£117.64	Dominated	0.00
Whooley + EPDS (cutoff 15)	0.8467	£82.34	£123.69	£22,313.88	70.50
Whooley + EPDS (cutoff 16)	0.8465	£78.28	£126.93		66.02

Notes: *ICERs calculated vs. next lower costing non-dominated or extendedly dominated strategy.
NCB is net consumption benefit. The cost-effective strategy at k=£30,000 is marked in **bold**.
ED= Extendedly Dominated

Appendix 6.5: Cost-effectiveness results using cross cohort measures for SDQ/RBS.

Appendix 6.5A Items in Both the SDQ and RBS Questionnaires

Item	SDQ Item	RBS Item	Responses
1	Restless, overactive, cannot stay still long	Very restless. Often running about or jumping up and down. Hardly ever still.	Doesn't (=0), somewhat (=1), or certainly applies (=2)
2	Often has temper tantrums	Irritable. Is quick to fly off the handle	Doesn't (=0), somewhat (=1), or certainly applies (=2)
3	Tends to play alone	Tends to do things on his/her own – rather solitary	Doesn't (=0), somewhat (=1), or certainly applies (=2)
5	Generally obedient*	Is often disobedient	Doesn't (=0), somewhat (=1), or certainly applies (=2)
6	Often seems worried	Often worried, worries about many things	Doesn't (=0), somewhat (=1), or certainly applies (=2)
7	Constantly fidgeting	Is squirmy or fidgety	Doesn't (=0), somewhat (=1), or certainly applies (=2)
8	Fights with or bullies other children	Bullies other children	Doesn't (=0), somewhat (=1), or certainly applies (=2)
9	Often unhappy	Often appears miserable, unhappy, tearful or distressed	Doesn't (=0), somewhat (=1), or certainly applies (=2)
10	Generally liked by other children*	Not much liked by other children	Doesn't (=0), somewhat (=1), or certainly applies (=2)
11	Easily distracted	Cannot settle to anything for more than a few moments	Doesn't (=0), somewhat (=1), or certainly applies (=2)
12	Picked on or bullied by other children	Frequently fights other children	Doesn't (=0), somewhat (=1), or certainly applies (=2)
13	Many fears, easily scared	Tends to be fearful or afraid of new things or new situations	Doesn't (=0), somewhat (=1), or certainly applies (=2)

Notes: Items used to generate the cross cohort measure.
Cross cohort measure generated by summing responses to produce a continuous variable from 0-26.
*Item is opposite, inverse scoring system applied e.g. responses are doesn't (=2), somewhat (=1), and certainly (=0) applies.

Appendix 6.5B: Mean Incremental Effects of Postnatal Depression Exposure on Child Development Outcomes at age 11

	Mean	95% Lower	95% Upper
GCA	-0.270	-.510	0.970)
SDQ Total Difficulties Cross Cohort Index	1.201	0.821	1.581

Notes: Incremental effect size and 95% confidence interval for children exposed to postnatal depression symptoms vs. no symptoms for the SDQ cross cohort index.
Results obtained from chapter three, growth curve model 1.
SDQ: Individuals n=11,804*;* Observations n=40,401*;* Exposed to PND n=1706
GCA: Individuals n=5,457; Observations n=19,186; Exposed to PND n=775

Appendix 6.5C: Mean Marginal Effects of Postnatal Depression Exposure on Lifetime Outcomes for RBS Cross Cohort Index

	QALYs	Healthcare Costs	Returns to Education*	Crime Costs
GCA	0.001 (0.001, 0.002)	-£13.56 (-£27.62, -£0.50)	£4,614.01 (£3,284, £5,947)	-£1.69 (-£2.79, -£0.60)
RBS Cross Cohort Index	-0.004 (-0.006, -0.001)	£87.70 (£22.99 £152.42)	-£212.40 (-£6,373, £6,012)	£6.81 (£1.14, £12.47)

Notes: Marginal effects of General Cognitive Ability and the Rutter Behavioural Scale cross cohort index on lifetime outcomes discounted at 3.5%. Each marginal effect is per unit increase in GCA and RBS. Results obtained from GLMs in chapter five
*Results were back transformed to the raw scale using Duan's smearing estimator.

Appendix 6.5D: Mean Incremental Lifetime Effects in Children Exposed to Postnatal Depression for SDQ/RBS Cross Cohort Index

	QALY s	Healthcare Costs	Returns to Education	Crime Costs
Postnatal depression symptoms (9 months)	-0.01	£108.99	-£1,500.88	£8.64

Notes: Incremental lifetime outcomes for children exposed to symptoms of postnatal depression vs children not exposed to symptoms of depression. Lifetime outcomes estimated indirectly using intermediate measures of GCA and the SDQ/ RBS *cross cohort* index. Lifetime outcomes are discounted at 3.5%.

Appendix 6.5E: Deterministic Cost-effectiveness Results for Cross Cohort SDQ /RBS Measure

Strategy	QALYs	Healthcare Costs	Education Costs	Health Centric Perspective ICERs*	Cross-Sectoral Results: NCB k=£30,000 v=£30,000
EPDS (cutoff 7)	0.844500	£192.51	£59.80	Dominated	-127.87
EPDS (cutoff 8)	0.845700	£180.10	£59.80	Dominated	-79.46
EPDS (cutoff 9)	0.846700	£164.86	£61.34	Dominated	-35.76
EPDS (cutoff 10)	0.847300	£156.48	£62.27	Dominated	-10.31
PHQ9 (cutoff 9/10)	0.847300	£150.55	£64.03	Dominated	-6.14
EPDS (cutoff 11)	0.847700	£144.17	£64.75	Dominated	11.52
EPDS (cutoff 12)	0.847900	£139.04	£65.83	Dominated	21.57
EPDS (cutoff 13)	0.847800	£138.76	£66.32	Dominated	18.36
EPDS (cutoff 14)	0.848000	£127.91	£69.67	Dominated	31.86
Whooley + EPDS (cutoff 7)	0.847700	£121.92	£60.99	Dominated	37.53
EPDS (cutoff 15)	0.847900	£118.04	£73.38	Dominated	35.02
Whooley + EPDS (cutoff 8)	0.848100	£117.58	£60.98	Dominated	53.88
EPDS (cutoff 16)	0.847800	£112.53	£75.42	Dominated	35.49
Whooley + EPDS (cutoff 9)	0.848400	£110.62	£62.45	Dominated	68.37
Whooley + EPDS (cutoff 10)	0.848500	£106.70	£63.34	Dominated	74.40
Whooely + PHQ9 (cutoff 9/10)	0.848500	£102.78	£65.00	Dominated	76.66
Whooley + EPDS (cutoff 11)	0.848608	£99.78	£65.69	£367,500.00	82.21
Whooley + EPDS (cutoff 12)	0.8486	£96.84	£66.72	**£27,714.29**	**83.88**
Whooley + EPDS (cutoff 13)	0.8485	£96.23	£67.18	ED	81.03
Whooley + EPDS (cutoff 14)	0.8483	£88.89	£70.37	Dominated	79.18
Standard Care	0.8456	£87.17	£70.27	Dominated	0.00
Whooley + EPDS (cutoff 15)	0.8481	£81.53	£73.89	Dominated	77.02
Whooley + EPDS (cutoff 16)	0.8479	£77.44	£75.83		73.17

Notes: *ICERs calculated vs. next lower costing non-dominated or extendedly dominated strategy.
NCB is net consumption benefit. The cost-effective strategy at k=£30,000 is marked in **bold**.
All costs and QALYs are discounted at 3.5%.
ED= Extendedly Dominated

References

ADALSTEINSSON, E. & TOUMI, M. 2013. Benefits of probabilistic sensitivity analysis–a review of NICE decisions. *Journal of market access & health policy,* 1**,** 21240.

ADES, A., LU, G. & CLAXTON, K. 2004. Expected value of sample information calculations in medical decision modeling. *Medical Decision Making,* 24**,** 207-227.

AKOBENG, A. 2005. Understanding randomised controlled trials. *Archives of disease in childhood,* 90**,** 840-844.

ALFONSO, V. C., FLANAGAN, D. P. & RADWAN, S. 2005. The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. *Contemporary intellectual assessment: Theories, tests, and***,** 185-202.

ANDRONIS, L., BARTON, P. & BRYAN, S. 2009. *Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making*, Prepress Projects Limited.

ANGELIS, A., KANAVOS, P. & MONTIBELLER, G. 2017. Resource Allocation and Priority Setting in Health Care: A Multi-criteria Decision Analysis Problem of Value? *Global Policy,* 8**,** 76-83.

ANISFELD, M. 2014. *Language development from birth to three*, Psychology Press.

ARKSEY, H. & O'MALLEY, L. 2005. Scoping studies: towards a methodological framework. *International journal of social research methodology,* 8**,** 19-32.

ARONSON, J. 2005. Biomarkers and surrogate endpoints. *British journal of clinical pharmacology,* 59**,** 491-494.

ARROW, K. J., CROPPER, M. L., GOLLIER, C., GROOM, B., HEAL, G. M., NEWELL, R. G., NORDHAUS, W. D., PINDYCK, R. S., PIZER, W. A. & PORTNEY, P. R. 2014. Should governments use a declining discount rate in project analysis? *Review of Environmental Economics and Policy,* 8**,** 145-163.

ASARIA, M. 2017. *Health care costs in the English NHS: reference tables for average annual NHS spend by age, sex and deprivation group*, Centre for Health Economics, The University of York.

ATKINS-BURNETT, S. & MEISELS, S. J. 2001. Measures of Socio-Emotional Development in Middle Childhood. Working Paper No. 2001-03. *National Center for Education Statistics*.

ATKINSON, M. 2015. Millennium Cohort Study-Interpreting the CANTAB Cognitive Measures. London, UK: Centre for Longitudinal Studies. *Institute of Education, University of London*.

ATTEMA, A. E., BROUWER, W. B. F. & CLAXTON, K. 2018. Discounting in Economic Evaluations. *PharmacoEconomics,* 36**,** 745-758.

BARBER, J. & THOMPSON, S. 2004. Multiple regression of cost data: use of generalised linear models. *Journal of health services research & policy,* 9**,** 197-204.

BARTON, G. R., BRIGGS, A. H. & FENWICK, E. A. 2008. Optimal cost-effectiveness decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI). *Value in Health,* 11**,** 886-897.

BATTISTA, C., EVANS, T. M., NGOON, T. J., CHEN, T., CHEN, L., KOCHALKA, J. & MENON, V. 2018. Mechanisms of interactive specialization and emergence of functional brain circuits supporting cognitive development in children. *npj Science of Learning,* 3**,** 1.

BECK, A. T. & ALFORD, B. A. 2009. *Depression: Causes and treatment*, University of Pennsylvania Press.

BECK, C. T. 1998. The effects of postpartum depression on child development: a meta-analysis. *Archives of psychiatric nursing,* 12**,** 12-20.

BELFIELD, C. R. & LEVIN, H. M. 2007. *The price we pay: Economic and social consequences of inadequate education*, Brookings Institution Press.

BELL, K. J. 2014. *Should women be screened for postnatal depression? Exploring the effects of undiagnosed maternal mental health problems on child development*. University of York.

BEN-SHLOMO, Y. & KUH, D. 2002. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of epidemiology,* 31**,** 285-293.
BERK, L. E. 2013. *Child Development*, Pearson Education.
BIESANZ, J. C., DEEB-SOSSA, N., PAPADAKIS, A. A., BOLLEN, K. A. & CURRAN, P. J. 2004. The role of coding time in estimating and interpreting growth curve models. *Psychological methods,* 9**,** 30.
BJORKLUND, D. F. & CAUSEY, K. B. 2017. *Children's thinking: Cognitive development and individual differences*, SAGE Publications.
BLACKORBY, C. & DONALDSON, D. 1990. A review article: The case against the use of the sum of compensating variations in cost-benefit analysis. *Canadian Journal of Economics***,** 471-494.
BLACKORBY, C., BOSSERT, W. & DONALDSON, D. 2002. Utilitarianism and the theory of justice. *Handbook of social choice and welfare,* 1**,** 543-596.
BOADWAY, R. W. 1974. The welfare foundations of cost-benefit analysis. *The Economic Journal,* 84**,** 926-939.
BOGIN, B. & SMITH, B. H. 1996. Evolution of the human life cycle. *American Journal of Human Biology: The Official Journal of the Human Biology Association,* 8**,** 703-716.
BOJKE, L., CLAXTON, K., SCULPHER, M. & PALMER, S. 2009. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value in Health,* 12**,** 739-749.
BOND, M., MEALING, S., ANDERSON, R., ELSTON, J., WEINER, G., TAYLOR, R. S., HOYLE, M., LIU, Z., PRICE, A. & STEIN, K. 2009. The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: a systematic review and economic model.
BOSANQUET, K., BAILEY, D., GILBODY, S., HARDEN, M., MANEA, L., NUTBROWN, S. & MCMILLAN, D. 2015. Diagnostic accuracy of the Whooley questions for the identification of depression: a diagnostic meta-analysis. *BMJ open,* 5**,** e008913.
BOTHWELL, L. E., GREENE, J. A., PODOLSKY, S. H. & JONES, D. S. 2016. Assessing the gold standard—lessons from the history of RCTs. Mass Medical Soc.
BOWERMAN, B. L., O'CONNELL, R. T. & KOEHLER, A. B. 2005. *Forecasting, Time Series, and Regression: An Applied Approach*, Thomson Brooks/Cole.
BOWLBY, J. 1978. Attachment theory and its therapeutic implications. *Adolescent Psychiatry,* 6**,** 5-33.
BOX, G. E. 1976. Science and statistics. *Journal of the American Statistical Association,* 71**,** 791-799.
BOYD, D. R., BEE, H. L. & JOHNSON, P. A. 2014. *Lifespan development*, Pearson.
BRADLEY, R. H. & CORWYN, R. F. 2002. Socioeconomic status and child development. *Annual review of psychology,* 53**,** 371-399.
BRAND, S. & PRICE, R. 2000. The economic and social costs of crime.
BRAND, S. R. & BRENNAN, P. A. 2009. Impact of antenatal and postpartum maternal mental illness: how are the children? *Clinical obstetrics and gynecology,* 52**,** 441-455.
BRAZIER, J., RATCLIFFE, J., SALOMAN, J. & TSUCHIYA, A. 2017. *Measuring and valuing health benefits for economic evaluation*, OXFORD university press.
BRENNAN, A. & AKEHURST, R. 2000. Modelling in health economic evaluation. *Pharmacoeconomics,* 17**,** 445-459.
BRENNAN, A., CHICK, S. E. & DAVIES, R. 2006. A taxonomy of model structures for economic evaluation of health technologies. *Health economics,* 15**,** 1295-1310.
BRENNAN, A., KHARROUBI, S., O'HAGAN, A. & CHILCOTT, J. 2007. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Medical Decision Making,* 27**,** 448-470.
BRENNAN, P. A., HAMMEN, C., ANDERSEN, M. J., BOR, W., NAJMAN, J. M. & WILLIAMS, G. M. 2000. Chronicity, severity, and timing of maternal depressive symptoms: relationships with child outcomes at age 5. *Developmental psychology,* 36**,** 759.

BRIGGS, A., GOEREE, R., BLACKHOUSE, G. & O'BRIEN, B. J. 2002. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Medical decision making,* 22**,** 290-308.
BRIGGS, A., CLARK, T., WOLSTENHOLME, J. & CLARKE, P. 2003. Missing.... presumed at random: cost-analysis of incomplete data. *Health economics,* 12**,** 377-392.
BRIGGS, A., SCULPHER, M. & CLAXTON, K. 2006. *Decision modelling for health economic evaluation*, OUP Oxford.
BRONFENBRENNER, U. 1977. Toward an experimental ecology of human development. *American psychologist,* 32**,** 513.
BRONFENBRENNER, U. 1992. *Ecological systems theory*, Jessica Kingsley Publishers.
BROTH, M. R., GOODMAN, S. H., HALL, C. & RAYNOR, L. C. 2004. Depressed and well mothers' emotion interpretation accuracy and the quality of mother—infant interaction. *Infancy,* 6**,** 37-55.
BROUWER, W. B. & KOOPMANSCHAP, M. A. 2000. On the economic foundations of CEA. Ladies and gentlemen, take your positions! *Journal of health economics,* 19**,** 439-459.
BROUWER, W. B., CULYER, A. J., VAN EXEL, N. J. A. & RUTTEN, F. F. 2008. Welfarism vs. extra-welfarism. *Journal of health economics,* 27**,** 325-338.
BRUCE, A., CHAMBERS, B., WRIGHT, J., BARRETT, B. T., BLOJ, M. & SHELDON, T. 2017. Visual acuity and early literacy at 6-7 years: A reduction in visual acuity is associated with decreased reading efficiency in school children. *Investigative Ophthalmology & Visual Science,* 58**,** 2367-2367.
BUNTIN, M. B. & ZASLAVSKY, A. M. 2004. Too much ado about two-part models and transformation?: Comparing methods of modeling Medicare expenditures. *Journal of Health Economics,* 23**,** 525-542.
BURNHAM, K. P. & ANDERSON, D. R. 2003. *Model selection and multimodel inference: a practical information-theoretic approach*, Springer Science & Business Media.
BUXBAUM, E. 2007. *Fundamentals of protein structure and function*, Springer.
BUXTON, M. J., DRUMMOND, M. F., VAN HOUT, B. A., PRINCE, R. L., SHELDON, T. A., SZUCS, T. & VRAY, M. 1997. Modelling in ecomomic evaluation: an unavoidable fact of life. *Health economics,* 6**,** 217-227.
BUXTON, M. J. 2006. Economic evaluation and decision making in the UK. *Pharmacoeconomics,* 24**,** 1133-1142.
CAMPBELL, F., PUNGELLO, E. P., BURCHINAL, M., KAINZ, K., PAN, Y., WASIK, B. H., BARBARIN, O. A., SPARLING, J. J. & RAMEY, C. T. 2012. Adult outcomes as a function of an early childhood educational program: an Abecedarian Project follow-up. *Developmental psychology,* 48**,** 1033.
CAMPBELL, F., CONTI, G., HECKMAN, J. J., MOON, S. H., PINTO, R., PUNGELLO, E. & PAN, Y. 2014. Early childhood investments substantially boost adult health. *Science,* 343**,** 1478-1485.
CARLSON, M. D. & MORRISON, R. S. 2009. Study design, precision, and validity in observational studies. *Journal of palliative medicine,* 12**,** 77-82.
CARLTON, J., KARNON, J., CZOSKI-MURRAY, C., SMITH, K. & MARR, J. 2008. The clinical effectiveness and cost-effectiveness of screening programmes for amblyopia and strabismus in children up to the age of 4–5 years: a systematic review and economic evaluation.
CASELLA, G. & BERGER, R. L. 2002. *Statistical inference*, Duxbury Pacific Grove, CA.
CAWLEY, J., HECKMAN, J. & VYTLACIL, E. 2001. Three observations on wages and measured cognitive ability. *Labour Economics,* 8**,** 419-442.
CHEN, C., CHOCK, D. P. & WINKLER, S. L. 1999. A simulation study of confounding in generalized linear models for air pollution epidemiology. *Environmental health perspectives,* 107**,** 217.

CIRELLI, C., POMPEIANO, M. & TONONI, G. 1996. Neuronal gene expression in the waking state: a role for the locus coeruleus. *Science,* 274**,** 1211-1215.
CLARKE, P., GRAY, A., BRIGGS, A., FARMER, A., FENN, P., STEVENS, R., MATTHEWS, D., STRATTON, I., HOLMAN, R. & GROUP, U. P. D. S. 2004. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia,* 47**,** 1747-1759.
CLAXTON, K. 1999. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of health economics,* 18**,** 341-364.
CLAXTON, K., SCULPHER, M. & DRUMMOND, M. 2002. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). *The Lancet,* 360**,** 711-715.
CLAXTON, K., SCULPHER, M., MCCABE, C., BRIGGS, A., AKEHURST, R., BUXTON, M., BRAZIER, J. & O'HAGAN, T. 2005. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health economics,* 14**,** 339-347.
CLAXTON, K., SCULPHER, M. & CULYER, T. 2007. Mark versus Luke? Appropriate methods for the evaluation of public health interventions. *Working Papers 031cherp,* Centre for Health Economics, University of York.
CLAXTON, K., WALKER, S., PALMER, S. & SCULPHER, M. 2010. Appropriate perspectives for health care decisions. *Working Papers 054cherp.* Centre for Health Economics, University of York.
CLAXTON, K., PAULDEN, M., GRAVELLE, H., BROUWER, W. & CULYER, A. J. 2011. Discounting and decision making in the economic evaluation of health-care technologies. *Health economics,* 20**,** 2-15.
CLAXTON, K., MARTIN, S., SOARES, M., RICE, N., SPACKMAN, E., HINDE, S., DEVLIN, N., SMITH, P. C. & SCULPHER, M. 2013. Methods for the estimation of the NICE cost effectiveness threshold. Centre for Health Economics, University of York.
COAST, J., SMITH, R. & LORGELLY, P. 2008. Should the capability approach be applied in health economics? *Health economics,* 17**,** 667-670.
COLQUHOUN, H. L., LEVAC, D., O'BRIEN, K. K., STRAUS, S., TRICCO, A. C., PERRIER, L., KASTNER, M. & MOHER, D. 2014. Scoping reviews: time for clarity in definition, methods, and reporting. *Journal of clinical epidemiology,* 67**,** 1291-1294.
CONNELLY, R. 2013. Millennium Cohort Study data note: Interpreting test scores. *London, UK.: Centre for Longitudinal Studies*.
COOKE, A., SMITH, D. & BOOTH, A. 2012. Beyond PICO the SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research,* 22**,** 1435-1443.
COOKSON, R. 2005. QALYs and the capability approach. *Health economics,* 14**,** 817-829.
COOKSON, R., COTTON-BARRETT, O., ADLER, M. D., ASARIA, M. & ORD, T. 2016. Years of Good Life Based on Income and Health: Re-Engineering Cost-Benefit Analysis to Examine Policy Impact on Wellbeing and Distributive Justice.
COOPER, N., SUTTON, A., ADES, A., PAISLEY, S., JONES, D. & MODELS', W. G. O. T. U. O. E. I. E. D. 2007. Use of evidence in economic decision models: practical issues and methodological challenges. *Health Economics,* 16**,** 1277-1286.
COOPER, N. J., SUTTON, A. J., ABRAMS, K. R., TURNER, D. & WAILOO, A. 2004. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health economics,* 13**,** 203-226.
CORBACHO, B. & PINTO-PRADES, J. L. 2012. Health economic decision-making: a comparison between UK and Spain. *British medical bulletin,* 103**,** 5-20.
CORNISH, A. M., MCMAHON, C. A., UNGERER, J. A., BARNETT, B., KOWALENKO, N. & TENNANT, C. 2006. Maternal depression and the experience of parenting in the second postnatal year. *Journal of Reproductive and Infant Psychology,* 24**,** 121-132.
COX, J. L., HOLDEN, J. M. & SAGOVSKY, R. 1987. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. *The British journal of psychiatry,* 150**,** 782-786.

CRICK, F. H. On protein synthesis. Symp Soc Exp Biol, 1958. 8.
CUIJPERS, P., BRÄNNMARK, J. G. & VAN STRATEN, A. 2008. Psychological treatment of postpartum depression: a meta-analysis. *Journal of clinical psychology,* 64**,** 103-118.
CUIJPERS, P., WEITZ, E., KARYOTAKI, E., GARBER, J. & ANDERSSON, G. 2015. The effects of psychological treatment of maternal depression on children and parental functioning: a meta-analysis. *European child & adolescent psychiatry,* 24**,** 237-245.
CULYER, T. 1991. The normative economics of health care finance and provision. *Providing Health Care: the Economics of Alternative Systems of Finance and Delivery*. Oxford University Press.
CUMMINGS, E. M. & DAVIES, P. T. 1994. Maternal depression and child development. *Journal of child psychology and psychiatry,* 35**,** 73-122.
CUNHA, F. & HECKMAN, J. 2007. The technology of skill formation. *American Economic Review,* 97**,** 31-47.
CUNHA, F. & HECKMAN, J. J. 2008. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources,* 43**,** 738-782.
CUNHA, F., HECKMAN, J. J. & SCHENNACH, S. M. 2010. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica,* 78**,** 883-931.
CURRAN, P. J., OBEIDAT, K. & LOSARDO, D. 2010. Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development,* 11**,** 121-136.
CURRAN, P. J., MCGINLEY, J. S., SERRANO, D. & BURFEIND, C. 2012. A multivariate growth curve model for three-level data. *APA handbook of research methods in psychology,* 3**,** 335-358.
CURRIE, J. 2001. Early childhood education programs. *Journal of Economic perspectives,* 15**,** 213-238.
DESMARAIS, B. A. & HARDEN, J. J. 2013. Testing for zero inflation in count models: Bias correction for the Vuong test. *The Stata Journal,* 13**,** 810-835.
DEX, S. & JOSHI, H. 2004. *Millennium Cohort Study first survey: A user's guide to initial findings*, Centre for Longitudinal Studies, Institute of Education, University of London.
DICKERSON, J., BIRD, P. K., MCEACHAN, R. R., PICKETT, K. E., WAIBLINGER, D., UPHOFF, E., MASON, D., BRYANT, M., BYWATER, T. & BOWYER-CRANE, C. 2016. Born in Bradford's Better Start: an experimental birth cohort study to evaluate the impact of early life interventions. BMC public health, 16, 711.
DOBSON, A. J. & BARNETT, A. 2008. *An introduction to generalized linear models*, CRC press.
DOLAN, P. & EDLIN, R. 2002. Is it really possible to build a bridge between cost-benefit analysis and cost-effectiveness analysis? *Journal of Health Economics,* 21**,** 827-843.
DOLL, R. & HILL, A. B. 1956. Lung cancer and other causes of death in relation to smoking. *British medical journal,* 2**,** 1071.
DONALDSON, C., MASON, H. & SHACKLEY, P. 2006. Contingent valuation in health care. *Chapters, in; The Elgar companion to health economics.*
ĐORĐIĆ, V., TUBIĆ, T. & JAKŠIĆ, D. 2016. The relationship between physical, motor, and intellectual development of preschool children. *Procedia-Social and Behavioral Sciences,* 233**,** 3-7.
DOWRICK, C., LEYDON, G.M., MCBRIDE, A., HOWE, A., BURGESS, H., CLARKE, P., MAISEY, S. & KENDRICK, T., 2009. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *Bmj*, *338*, p.b663.
DRUMMOND, M. F., SCULPHER, M. J., CLAXTON, K., STODDART, G. L. & TORRANCE, G. W. 2015. *Methods for the economic evaluation of health care programmes*, Oxford university press.
DYKXHOORN, D. M., NOVINA, C. D. & SHARP, P. A. 2003. Killing the messenger: short RNAs that silence gene expression. *Nature reviews Molecular cell biology,* 4**,** 457.
EICHLER, H. G., KONG, S. X., GERTH, W. C., MAVROS, P. & JÖNSSON, B. 2004. Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? *Value in health,* 7**,** 518-528.
ELLIOTT, J. & SHEPHERD, P. 2006. Cohort profile: 1970 British birth cohort (BCS70). *International journal of epidemiology,* 35**,** 836-843.

EVANS, G. W. 2006. Child development and the physical environment. *Annu. Rev. Psychol.,* 57**,** 423-451.
FALKNER, B., DANIELS, S. R., FLYNN, J. T., GIDDING, S., GREEN, L. A., INGELFINGER, J. R., LAUER, R. M., MORGENSTERN, B. Z., PORTMAN, R. J. & PRINEAS, R. J. 2004. The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents. *Pediatrics,* 114**,** 555-576.
FARRAR, D. E. & GLAUBER, R. R. 1967. Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics***,** 92-107.
FEDAK, K. M., BERNAL, A., CAPSHAW, Z. A. & GROSS, S. 2015. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerging themes in epidemiology*, 12, 14
FEIWEL, G. R. 2016. *Arrow and the Foundations of the Theory of Economic Policy*, Springer.
FELDMAN, A. M. 1989. Arrow's Impossibility Theorem. In: Welfare Economics and Social Choice Theory. *Springer, Boston, MA*.
FENWICK, E., CLAXTON, K. & SCULPHER, M. 2001. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health economics,* 10**,** 779-787.
FERGUSON, K. T., CASSELLS, R. C., MACALLISTER, J. W. & EVANS, G. W. 2013. The physical environment and child development: An international review. *International Journal of Psychology,* 48**,** 437-468.
FIEST, K. M., JETTE, N., QUAN, H., GERMAINE-SMITH, C. S., METCALFE, A., PATTEN, S. B. & BECK, C. A. 2014. Systematic review and assessment of validated case definitions for depression in administrative data. *BMC psychiatry,* 14**,** 289.
FINGERMAN, K. L., FINGERMAN, K. L., BERG, C., SMITH, J. & ANTONUCCI, T. C. 2011. *Handbook of Life-Span Development*, Springer Publishing Company.
FITELSON, E., KIM, S., BAKER, A. S. & LEIGHT, K. 2011. Treatment of postpartum depression: clinical, psychological and pharmacological options. *International journal of women's health,* 3**,** 1.
FLEMING, T. R. & DEMETS, D. L. 1996. Surrogate end points in clinical trials: are we being misled? *Annals of internal medicine,* 125**,** 605-613.
FLEMING, T. R. & POWERS, J. H. 2012. Biomarkers and surrogate endpoints in clinical trials. *Statistics in medicine,* 31**,** 2973-2984.
FLORA, G., GUPTA, D. & TIWARI, A. 2012. Toxicity of lead: a review with recent updates. *Interdisciplinary toxicology,* 5**,** 47-58.
FLOURI, E., MIDOUHAS, E. & JOSHI, H. 2014. Family poverty and trajectories of children's emotional and behavioural problems: the moderating roles of self-regulation and verbal cognitive ability. *Journal of abnormal child psychology,* 42**,** 1043-1056.
FRISELL, T., PAWITAN, Y. & LÅNGSTRÖM, N. 2012. Is the association between general cognitive ability and violent crime caused by family-level confounders? *PloS one,* 7**,** e41783.
FURBER, G., SEGAL, L., LEACH, M. & COCKS, J. 2014. Mapping scores from the Strengths and Difficulties Questionnaire (SDQ) to preference-based utility values. *Quality of Life Research,* 23**,** 403-411.
GAINS, F., 2015. Metro mayors: devolution, democracy and the importance of getting the 'Devo Manc'design right. *Representation*, *51*(4), pp.425-437.
GARNETT, G. P., COUSENS, S., HALLETT, T. B., STEKETEE, R. & WALKER, N. 2011. Mathematical models in the evaluation of health programmes. *The Lancet,* 378**,** 515-525.
GAUVAIN, P. M. & PARKE, R. D. 2008. *Child Psychology: A Contemporary View Point*, McGraw-Hill Education.
GILBERT-BARNESS, E. 2010. Teratogenic causes of malformations. *Annals of Clinical & Laboratory Science,* 40**,** 99-114.
GILBERT, R., LOGAN, S., MOYER, V. A. & ELLIOTT, E. J. 2001. Assessing diagnostic and screening tests. *The Western journal of medicine,* 175**,** 37.

GILBODY, S.M., HOUSE, A.O. & SHELDON, T.A., 2001. Routinely administered questionnaires for depression and anxiety: systematic review. *Bmj*, *322*(7283), pp.406-409.

GJERDINGEN, D., CROW, S., MCGOVERN, P., MINER, M. & CENTER, B. 2009. Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. *The Annals of Family Medicine,* 7**,** 63-70.

GLANVILLE, J., LEFEBVRE, C. & WRIGHT, K. 2008. *The InterTASC Information Specialists' Sub-Group; 2008 [updated 2018 February 26; cited 2018]* [Online]. Available: https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home.

GLEWWE, P. 2002. Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of economic literature,* 40**,** 436-482.

GOLDBERG, D. 2014. The value of screening in patient populations with high prevalence of a disorder. *BMC medicine,* 12**,** 14.

GOODMAN, J. H. 2004. Postpartum depression beyond the early postpartum period. *Journal of Obstetric, Gynecologic, & Neonatal Nursing,* 33**,** 410-420.

GOODMAN, R. 1997. The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry,* 38**,** 581-586.

GOODMAN, R. 2001. Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry,* 40**,** 1337-1345.

GOODMAN, S. H. & GARBER, J. 2017. Evidence-Based Interventions for Depressed Mothers and Their Young Children. *Child Development,* 88**,** 368-377.

GOTTLIEB, G. 1998. Normally occurring environmental and behavioral influences on gene activity: from central dogma to probabilistic epigenesis. *Psychological review,* 105**,** 792.

GOYAL, D., GAY, C. & LEE, K. A. 2010. How much does low socioeconomic status increase the risk of prenatal and postpartum depressive symptoms in first-time mothers? *Women's Health Issues,* 20**,** 96-104.

GRACE, S. L., EVINDAR, A. & STEWART, D. 2003. The effect of postpartum depression on child cognitive development and behavior: a review and critical analysis of the literature. *Archives of women's mental health,* 6**,** 263-274.

GREENHALGH, T. 1997. How to read a paper. Papers that report diagnostic or screening tests. *BMJ: British Medical Journal,* 315**,** 540.

GREER, S.L., 2016. Devolution and health in the UK: policy and its lessons since 1998. *British medical bulletin*, *118*(1), p.16.

GROSSMAN, M. 1972. On the concept of health capital and the demand for health. *Journal of Political economy,* 80**,** 223-255.

HALFON, N., INKELAS, M. & HOCHSTEIN, M. 2000. The health development organization: An organizational approach to achieving child health development. *The Milbank Quarterly,* 78**,** 447-497.

HALFON, N. & HOCHSTEIN, M. 2002. Life course health development: an integrated framework for developing health, policy, and research. *The Milbank Quarterly,* 80**,** 433-479.

HALFON, N., LARSON, K., LU, M., TULLIS, E. & RUSS, S. 2014. Lifecourse health development: past, present and future. *Maternal and child health journal,* 18**,** 344-365.

HALLE, T. G. & DARLING-CHURCHILL, K. E. 2016. Review of measures of social and emotional development. *Journal of Applied Developmental Psychology,* 45**,** 8-18.

HAMMEN, C. & BRENNAN, P. A. 2003. Severity, chronicity, and timing of maternal depression and risk for adolescent offspring diagnoses in a community sample. *Archives of general psychiatry,* 60**,** 253-258.

HAMMER, D., MELHUISH, E. & HOWARD, S. J. 2017. Do aspects of social, emotional and behavioural development in the pre-school period predict later cognitive and academic attainment? *Australian Journal of Education,* 61**,** 270-287.

HAMMOND, P. J. 1991. Interpersonal comparisons of utility: Why and how they are and should be made. *Interpersonal comparisons of well-being***,** 200-254.

HANSON, M. A. & GLUCKMAN, P. D. 2008. Developmental origins of health and disease: new insights. *Basic & clinical pharmacology & toxicology,* 102**,** 90-93.
HANUSHEK, E. A. & WOESSMANN, L. 2008. The role of cognitive skills in economic development. *Journal of economic literature,* 46**,** 607-68.
HAWKINS, N., RICHARDSON, G., SUTTON, A. J., COOPER, N. J., GRIFFITHS, C., ROGERS, A. & BOWER, P. 2012. Surrogates, meta-analysis and cost-effectiveness modelling: a combined analytic approach. *Health economics,* 21**,** 742-756.
HAY, D. F., PAWLBY, S., WATERS, C. S. & SHARP, D. 2008. Antepartum and postpartum exposure to maternal depression: different effects on different adolescent outcomes. *Journal of Child Psychology and Psychiatry,* 49**,** 1079-1088.
HEALEY, A., KNAPP, M. & FARRINGTON, D. P. 2004. Adult labour market implications of antisocial behaviour in childhood and adolescence: findings from a UK longitudinal study. *Applied Economics,* 36**,** 93-105.
HEATH, A., MANOLOPOULOU, I. & BAIO, G. 2016. Estimating the expected value of partial perfect information in health economic evaluations using integrated nested Laplace approximation. *Statistics in medicine,* 35**,** 4264-4280.
HECKMAN, J. J. 2006. Skill formation and the economics of investing in disadvantaged children. *Science,* 312**,** 1900-1902.
HECKMAN, J. J. 2008. The case for investing in disadvantaged young children. *CESifo DICE Report,* 6**,** 3-8.
HECKMAN, J. J., MOON, S. H., PINTO, R., SAVELYEV, P. A. & YAVITZ, A. 2010. The rate of return to the HighScope Perry Preschool Program. *Journal of public Economics,* 94**,** 114-128.
HEINDEL, J. J. & VANDENBERG, L. N. 2015. Developmental origins of health and disease: a paradigm for understanding disease etiology and prevention. *Current opinion in pediatrics,* 27**,** 248.
HERRNSTEIN, R. J. & MURRAY, C. 2010. *Bell curve: Intelligence and class structure in American life*, Simon and Schuster.
HERTZMAN, C., POWER, C., MATTHEWS, S. & MANOR, O. 2001. Using an interactive framework of society and lifecourse to explain self-rated health in early adulthood. *Social Science and Medicine,* 53**,** 1575-1585.
HEWITT, C., GILBODY, S., BREALEY, S., PAULDEN, M., PALMER, S., MANN, R., GREEN, J., MORRELL, J., BARKHAM, M. & LIGHT, K. 2009. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis.
HEX, N., BARTLETT, C., WRIGHT, D., TAYLOR, M. & VARLEY, D. 2012. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabetic Medicine,* 29**,** 855-862.
HEYLAND, D.K., GAFNI, A., KERNERMAN, P., KEENAN, S. & CHALFIN, D., 1999. How to use the results of an economic evaluation. *Critical care medicine*, *27*(6), pp.1195-1202.
HEYWOOD, J., JOHNSON, J. & BROWN, M. 2015. British Cohort Study User Guide to the Response and Deaths Datasets. *Institute of Education, University of London*.
HILL, V. 2005. Through the past darkly: a review of the British ability scales second edition. *Child and Adolescent Mental Health,* 10**,** 87-98.
HINDE, S., WIYANI, A., GRIFFIN, S. AND WALKER, S., 2017. Who bears the cost of NICE public health recommendations? *British medical bulletin*, *124*(1), pp.113-120.
HINTERMAIR, M. 2006. Parental resources, parental stress, and socioemotional development of deaf and hard of hearing children. *The Journal of Deaf Studies and Deaf Education,* 11**,** 493-513.
HM R&C (HER MAJESTY'S REVENUE AND CUSTOMS). 2017. *Distribution of median and mean income and tax by age range and gender* [Online]. Gov.uk [Internet]. Available: https://www.gov.uk/government/statistics/distribution-of-median-and-mean-income-and-tax-by-age-range-and-gender-2010-to-2011.

HM TREASURY (HER MAJESTY'S TREASURY). 2018. The Green Book: Central Government Guidance on Appraisal and Evaluation. 2018. *London: HM Treasury*.
HOLLINGWORTH, W., HAWKINS, J., LAWLOR, D., BROWN, M., MARSH, T. & KIPPING, R. 2012. Economic evaluation of lifestyle interventions to treat overweight or obesity in children. *International Journal of Obesity*, 36, 559.
HOME OFFICE. 1997. *Digest 4. Chapter 3- Offending and Offenders.* [Online]. Available: http://webarchive.nationalarchives.gov.uk/20110220155237/http://rds.homeoffice.gov.uk/rds/digest4/chapter3.pdf [Accessed 08/02/2018.
HUMMEL, S., CHILCOTT, J., RAWDIN, A. & STRONG, M. 2011. *Economic outcomes of early years programmes and interventions designed to promote cognitive, social and emotional development among vulnerable children and families. ScHARR Public Health Evidence Report 9.5* [Online]. Available: https://www.sheffield.ac.uk/scharr/sections/ph/series [Accessed 02/08/2018.
IHGSC (INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM), 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860.
JACKSON, C., STEVENS, J., REN, S., LATIMER, N., BOJKE, L., MANCA, A. & SHARPLES, L. 2016. Extrapolating Survival from Randomized Trials Using External Data A Review of Methods. *Medical Decision Making***,** 0272989X16639900.
JAKOBSEN, J. C., GLUUD, C., WETTERSLEV, J. & WINKEL, P. 2017. When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology*, 17, 162.
JENSEN, A. R. 1998. *The g factor: The science of mental ability*, Praeger Westport, CT.
JOHANNESSON, M. 1996. *Theory and methods of economic evaluation of health care*, Springer Science & Business Media.
JOHNSON, J. & SETAKIS, E. 2015. Millennium Cohort Study MCS4: A guide to the Linked Education Administrative Datasets. *Institute of Education, University of London*.
JONES, A. M. & O'DONNELL, O. 2002. *Econometric analysis of health data*, Wiley Online Library.
JONES, A. M. 2010. Models for health care. *Health, Econometrics and Data Group (HEDG) Working Papers*.
JONES, E. M. & SCHOON, I. 2008. Child cognition and behaviour. *Millennium Cohort Study Third Survey: A user's guide to initial findings***,** 118-144.
KAPLAN, D. 2018. Causal Inference for Observational Studies. *The Journal of Infectious Diseases, jiy392-jiy392*.
KEENAN, T. & EVANS, S. 2009. *An introduction to child development*, Sage.
KELLY, Y., BECARES, L. & NAZROO, J. 2012. Associations between maternal experiences of racism and early child health and development: findings from the UK Millennium Cohort Study. *J Epidemiol Community Health***,** jech-2011-200814.
KENDRICK, T., DOWRICK, C., MCBRIDE, A., HOWE, A., CLARKE, P., MAISEY, S., MOORE, M. & SMITH, P.W., 2009. Management of depression in UK general practice in relation to scores on depression severity questionnaires: analysis of medical record data. *Bmj*, *338*, p.b750.
KENNEDY, C. A. 2002. Revealed preference valuation compared to contingent valuation: radon-induced lung cancer prevention. *Health Economics,* 11**,** 585-598.
KESSLER, D., BENNEWITH, O., LEWIS, G. & SHARP, D. 2002. Detection of depression and anxiety in primary care: follow up study. *Bmj,* 325**,** 1016-1017.
KINGSTON, D. & TOUGH, S. 2014. Prenatal and postnatal maternal mental health and school-age child development: a systematic review. *Maternal and child health journal,* 18**,** 1728-1741.
KNUDSEN, E. I. 2004. Sensitive periods in the development of the brain and behavior. *Journal of cognitive neuroscience,* 16**,** 1412-1425.
KOONIN, E. V. 2012. Does the central dogma still stand? *Biology direct,* 7**,** 27.
KROENKE, K., SPITZER, R. L. & WILLIAMS, J. B. 2001. The phq-9. *Journal of general internal medicine,* 16**,** 606-613.

KROTKIEWSKI, M., MANDROUKAS, K., SJÖSTRÖM, L., SULLIVAN, L., WETTERQVIST, H. & BJÖRNTORP, P. 1979. Effects of long-term physical training on body fat, metabolism, and blood pressure in obesity. *Metabolism-Clinical and Experimental,* 28**,** 650-658.
KWOK, O.-M., UNDERHILL, A. T., BERRY, J. W., LUO, W., ELLIOTT, T. R. & YOON, M. 2008. Analyzing longitudinal data with multilevel models: An example with individuals living with lower extremity intra-articular fractures. *Rehabilitation psychology,* 53**,** 370.
LALKHEN, A. G. & MCCLUSKEY, A. 2008. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain,* 8**,** 221-223.
LAYARD, R., CLARK, A. E., CORNAGLIA, F., POWDTHAVEE, N. & VERNOIT, J. 2014. What predicts a successful life? A life-course model of well-being. *The Economic Journal,* 124.
LEAHY-WARREN, P. & MCCARTHY, G. 2007. Postnatal depression: prevalence, mothers' perspectives, and treatments. *Archives of psychiatric nursing,* 21**,** 91-100.
LEAL, J., WORDSWORTH, S., LEGOOD, R. & BLAIR, E. 2007. Eliciting expert opinion for economic models: an applied example. *Value in Health,* 10**,** 195-203.
LEIS, J. A., HERON, J., STUART, E. A. & MENDELSON, T. 2014. Associations between maternal mental health and child emotional and behavioral problems: does prenatal mental health matter? *Journal of abnormal child psychology,* 42**,** 161-171.
LEREYA, S. T., COPELAND, W. E., COSTELLO, E. J. & WOLKE, D. 2015. Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries. *The Lancet Psychiatry,* 2**,** 524-531.
LERNER, R. M. 2006. Developmental science, developmental systems, and contemporary theories of human development. *Handbook of child psychology*.
LERNER, R. M. & OVERTON, W. F. 2010. *The Handbook of Life-Span Development, Volume 1: Cognition, Biology, and Methods*, Wiley.
LERNER, R. M. & BUSCH-ROSSNAGEL, N. A. 2013. *Individuals as producers of their development: A life-span perspective*, Elsevier.
LEVAC, D., COLQUHOUN, H. & O'BRIEN, K. K. 2010. Scoping studies: advancing the methodology. *Implementation Science,* 5**,** 1.
LEVIN, H. M. & BELFIELD, C. 2015. Guiding the development and use of cost-effectiveness analysis in education. *Journal of Research on Educational Effectiveness,* 8**,** 400-418.
LEVIN, H. M., MCEWAN, P. J., BELFIELD, C., BOWDEN, A. B. & SHAND, R. 2017. *Economic Evaluation in Education: Cost-Effectiveness and Benefit-Cost Analysis*, SAGE Publications.
LIGHTFOOT, C. 2013. *The development of children,* New York, Worth.
LLOYD, A., SCHMIEDER, C. & MARCHANT, N. 2003. Financial and health costs of uncontrolled blood pressure in the United Kingdom. *Pharmacoeconomics,* 21**,** 33-41.
LOMAS, J., ASARIA, M., BOJKE, L., GALE, C. P., RICHARDSON, G. & WALKER, S. 2018. Which costs matter? Costs included in economic evaluation and their impact on decision uncertainty for stable coronary artery disease. *PharmacoEconomics-open***,** 1-11.
LONG, K. 2009. Unemployment durations: evidence from the British Household Panel Survey. *Economic & Labour Market Review,* 3**,** 48-54.
LUENGO-FERNANDEZ, R., LEAL, J. & GRAY, A. 2012. UK research expenditure on dementia, heart disease, stroke and cancer: are levels of spending related to disease burden? *European journal of Neurology,* 19**,** 149-154.
MACHINA, M. J. 1987. Choice under uncertainty: Problems solved and unsolved. *Journal of Economic Perspectives,* 1**,** 121-154.
MAKRIDAKIS, S., WHEELWRIGHT, S. C. & HYNDMAN, R. J. 2008. *Forecasting methods and applications, 3rd ed*, Wiley India Pvt. Limited.
MANN, R., ADAMSON, J. & GILBODY, S. M. 2012. Diagnostic accuracy of case-finding questions to identify perinatal depression. *Canadian Medical Association Journal,* 184**,** E424-E430.
MARTINEZ, R., SOLIZ, P., CAIXETA, R. & ORDUNEZ, P., 2019. Reflection on modern methods: years of life lost due to premature mortality—a versatile and comprehensive

measure for monitoring non-communicable disease mortality. *International journal of epidemiology*, *1*, p.10.

MASTEN, A. S. & REED, M.-G. J. 2002. Resilience in development. *Handbook of positive psychology*, 74-88.

MASTEN, A. S. & CICCHETTI, D. 2010. Developmental cascades. *Development and psychopathology,* 22**,** 491-495.

MCCOACH, D. B. & KANISKAN, B. 2010. Using time-varying covariates in multilevel growth models. *Frontiers in psychology,* 1.

MCCULLAGH, P. 1984. Generalized linear models. *European Journal of Operational Research,* 16**,** 285-292.

MCKINLAY, R. J., WILCZYNSKI, N. L. & HAYNES, R. B. 2006. Optimal search strategies for detecting cost and economic studies in EMBASE. *BMC Health Services Research,* 6**,** 67.

MCMUNN, A., KELLY, Y., CABLE, N. & BARTLEY, M. 2011. Maternal employment and child socio-emotional behaviour in the UK: longitudinal evidence from the UK Millennium Cohort Study. *Journal of Epidemiology & Community Health*, jech. 2010.109553.

MCPHERSON, K., MARSH, T., BROWN M. 2007. Tackling Obesities: Future Choices–Modelling Future Trends in Obesity and the Impact on Health 2nd Edition. *UK Government Foresight*.

MCRAE, J. F., CLAYTON, S., FITZGERALD, T. W., KAPLANIS, J., PRIGMORE, E., RAJAN, D., SIFRIM, A., AITKEN, S., AKAWI, N. & ALVI, M. 2017. Prevalence and architecture of de novo mutations in developmental disorders. *Nature,* 542**,** 433.

MEYER, P. A., BROWN, M. J. & FALK, H. 2008. Global approach to reducing lead exposure and poisoning. *Mutation research/reviews in mutation research,* 659**,** 166-175.

MILGROM, J., GEMMILL, A.W., ERICKSEN, J., BURROWS, G., BUIST, A. & REECE, J., 2015. Treatment of postnatal depression with cognitive behavioural therapy, sertraline and combination therapy: a randomised controlled trial. *Australian & New Zealand Journal of Psychiatry*, *49*(3), pp.236-245.

MITCHELL, P.M., AL-JANABI, H., BYFORD, S., KUYKEN, W., RICHARDSON, J., IEZZI, A. & COAST, J., 2017. Assessing the validity of the ICECAP-A capability measure for adults with depression. *BMC psychiatry*, *17*(1), p.46.

MOELLER, M. P. 2000. Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics,* 106**,** e43-e43.

MOLENAAR, P. C., HUIZENGA, H. M. & NESSELROADE, J. R. 2003. The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. *Understanding human development.* Springer.

MONIQUE VAN LONDEN, W., JUFFER, F. & VAN IJZENDOORN, M. H. 2007. Attachment, cognitive, and motor development in adopted children: Short-term outcomes after international adoption. *Journal of Pediatric Psychology,* 32**,** 1249-1258.

MORIZOT, J. & KAZEMIAN, L. 2014. *The development of criminal and antisocial behavior*, Springer.

MORRIS, S., DEVLIN, N. & PARKIN, D. 2007. *Economic analysis in health care*, John Wiley & Sons.

MOSTAFA, T. & WIGGINS, D., 2014. Handling attrition and non-response in the 1970 British Cohort Study.

MUENNIG, P., ROBERTSON, D., JOHNSON, G., CAMPBELL, F., PUNGELLO, E. P. & NEIDELL, M. 2011. The effect of an early education program on adult health: the Carolina Abecedarian Project randomized controlled trial. *American Journal of Public Health,* 101**,** 512-516.

MURAD, M. H., KATABI, A., BENKHADRA, R. & MONTORI, V. M. 2018. External validity, generalisability, applicability and directness: a brief primer. *Evidence-based medicine*, 23, 17-19.

MURRAY, L., ARTECHE, A., FEARON, P., HALLIGAN, S., GOODYER, I. & COOPER, P. 2011. Maternal postnatal depression and the development of depression in offspring up to 16 years of age. *Journal of the American Academy of Child & Adolescent Psychiatry,* 50**,** 460-470.
MURRAY, L. & COOPER, P. J. 1997. Postpartum depression and child development. *Psychological medicine,* 27**,** 253-260.
NALDINI, L. 2015. Gene therapy returns to centre stage. *Nature,* 526**,** 351.
NAVERŠNIK, K. & ROJNIK, K. 2012. Handling input correlations in pharmacoeconomic models. *Value in Health,* 15**,** 540-549.
NEWMAN, B. M. & NEWMAN, P. R. 2017. *Development through life: A psychosocial approach*, Cengage Learning.
NEWTON, J. N., BRIGGS, A. D., MURRAY, C. J., DICKER, D., FOREMAN, K. J., WANG, H., NAGHAVI, M., FOROUZANFAR, M. H., OHNO, S. L. & BARBER, R. M. 2015. Changes in health in England, with analysis by English regions and areas of deprivation, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet,* 386**,** 2257-2274.
NHS EXECUTIVE (NATIONAL HEALTH SERVICE EXECUTIVE REPORT). 1996. *Burdens of disease: a discussion document,* The British Library, Whetherby, NHS Executive.
NICE (NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE), 2013. National Institute for Health and Care Excellence Guide to the methods of technology appraisal 2013. NICE guidelines (PMG9).
NICE (NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE), 2018. *Antenatal and postnatal mental health: clinical management and service guidance. NICE guideline [CG192]* [Online]. Available: https://www.nice.org.uk/guidance/cg192 2018].
NOBLE JR, J. H. 2006. Meta-analysis: methods, strengths, weaknesses, and political uses. *Journal of Laboratory and Clinical Medicine,* 147**,** 7-20.
NUSSBAUM, M. 2003. Capabilities as fundamental entitlements: Sen and social justice. *Feminist economics,* 9**,** 33-59.
NUTTALL, A.K., FROYEN, L.C., SKIBBE, L.E. & BOWLES, R.P., 2019. Maternal and Paternal Depressive Symptoms, Home Learning Environment, and Children's Early Literacy. *Child Psychiatry & Human Development*, pp.1-11.
O'FLAHERTY, J. & PHILLIPS, C. 2015. The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education,* 25**,** 85-95.
OLSEN, J. A. & DONALDSON, C. 1998. Helicopters, hearts and hips: using willingness to pay to set priorities for public sector health care programmes. *Social Science & Medicine,* 46**,** 1-12.
ONS (OFFICE FOR NATIONAL STATISTICS), 2017. *Vital statistics: population and health reference tables* [Online]. London: Office for National Statistics. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/vitalstatisticspopulationandhealthreferencetables.
ONS (OFFICE FOR NATIONAL STATISTICS), 2018a. *Birth characteristics in England and Wales: 2016* [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthcharacteristicsinenglandandwales/2016.
ONS (OFFICE FOR NATIONAL STATISTICS), 2018b. *Consumer Price Inflation time series dataset (MM23)* [Online]. London: Office for National Statistics Available: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/l55o/mm23.
OSTER, G., THOMPSON, D., EDELSBERG, J., BIRD, A. P. & COLDITZ, G. A. 1999. Lifetime health and economic benefits of weight loss among obese persons. *American Journal of Public Health,* 89**,** 1536-1542.
OVERTON, W. F. 2003. Development across the life span. *Handbook of psychology*.
OVERTON, W. F. 2010a. Developmental Psychology. *The Corsini Encyclopedia of Psychology*. John Wiley & Sons, Inc.
OVERTON, W. F. 2010b. Life-Span Development. *The handbook of life-span development*.

PALMER, S. & RAFTERY, J. 1999. Economics notes: Opportunity cost. *BMJ: British Medical Journal,* 318**,** 1551.
PARSONS, S. 2014. Childhood Cognition in the 1970 British Cohort Study. Data Note. *Institute of Education, University of London*.
PAULDEN, M., PALMER, S., HEWITT, C. & GILBODY, S. 2009. Screening for postnatal depression in primary care: cost effectiveness analysis. *Bmj,* 339**,** b5203.
PEASGOOD, T., BRAZIER, J.E., MUKURIA, C. AND ROWEN, D., 2014. A conceptual comparison of well-being measures used in the UK.
PHAM, M. T., RAJIĆ, A., GREIG, J. D., SARGEANT, J. M., PAPADOPOULOS, A. & MCEWEN, S. A. 2014. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Research synthesis methods,* 5**,** 371-385.
PIGLIUCCI, M., MÜLLER, G. B., EVOLUTION, K. L. I. F. & RESEARCH, C. 2010. *Evolution, the Extended Synthesis*, MIT Press.
PLEWIS, I., CALDERWOOD, L., HAWKES, D., HUGHES, G. & JOSHI, H. 2007. Millennium Cohort Study: technical report on sampling. *London: Centre for Longitudinal Studies*.
POLIN, R. A., ABMAN, S. H., ROWITCH, D. & BENITZ, W. E. 2016. *Fetal and Neonatal Physiology E-Book*, Elsevier Health Sciences.
POWER, C. & HERTZMAN, C. 1997. Social and biological pathways linking early life and adult disease. *British medical bulletin,* 53**,** 210-221.
POWER, C. & ELLIOTT, J. 2006. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol,* 35**,** 34-41.
PREGIBON, D. 1980. Goodness of link tests for generalized linear models. Applied statistics, 15-14.
PRENTICE, R. L. 1989. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine,* 8**,** 431-440.
PROCHASKA, J. J., SUNG, H. Y., MAX, W., SHI, Y. & ONG, M. 2012. Validity study of the K6 scale as a measure of moderate mental distress based on mental health treatment need and utilization. *International journal of methods in psychiatric research,* 21**,** 88-97.
RABINER, D. L., GODWIN, J. & DODGE, K. A. 2016. Predicting academic achievement and attainment: the contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review,* 45**,** 250-267.
RAFTERY, J. 2014. Health economic evaluation in England. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen,* 108**,** 367-374.
RAMCHANDANI, P., STEIN, A., EVANS, J., O'CONNOR, T.G. & ALSPAC STUDY TEAM, 2005. Paternal depression in the postnatal period and child development: a prospective population study. *The Lancet*, *365*(9478), pp.2201-2205.
RELTON, C., TORGERSON, D., O'CATHAIN, A. & NICHOLL, J. 2010. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. *Bmj,* 340**,** c1066.
RICE, D. & BARONE JR, S. 2000. Critical periods of vulnerability for the developing nervous system: evidence from humans and animal models. *Environmental health perspectives,* 108**,** 511.
RICHARDSON, G. A. 2007. *The cost-effectiveness of interventions to support self care.* University of York.
RIDE, J. 2018. Setting the Boundaries for Economic Evaluation: Investigating Time Horizon and Family Effects in the Case of Postnatal Depression. *Value in Health,* 21**,** 573-580.
ROGERS, C. E., ANDERSON, P. J., THOMPSON, D. K., KIDOKORO, H., WALLENDORF, M., TREYVAUD, K., ROBERTS, G., DOYLE, L. W., NEIL, J. J. & INDER, T. E. 2012. Regional cerebral development at term relates to school-age social–emotional development in very preterm children. *Journal of the American Academy of Child & Adolescent Psychiatry,* 51**,** 181-191.
ROGOFF, B. 2003. *The cultural nature of human development*, Oxford University Press.
ROYSTON, P. 2004. Multiple imputation of missing values. *Stata journal,* 4**,** 227-41.
RUBIN, D. B. 2004. *Multiple imputation for nonresponse in surveys*, John Wiley & Sons.

RUDISILL, C., CHARLTON, J., BOOTH, H. & GULLIFORD, M. 2016. Are healthcare costs from obesity associated with body mass index, comorbidity or depression? Cohort study using electronic health records. *Clinical obesity,* 6**,** 225-231.
RUTTER, M. 1998. Developmental catch-up, and deficit, following adoption after severe global early privation. *The Journal of Child Psychology and Psychiatry and Allied Disciplines,* 39**,** 465-476.
SAMPSON, R. J. & LAUB, J. H. 2005. A life-course view of the development of crime. *The Annals of the American Academy of Political and Social Science,* 602**,** 12-45.
SANGER, C., ILES, J. E., ANDREW, C. S. & RAMCHANDANI, P. G. 2015. Associations between postnatal maternal depression and psychological outcomes in adolescent offspring: a systematic review. *Archives of women's mental health,* 18**,** 147-162.
SANTROCK, J. W. 2003. *Child Development: An Introduction*, McGraw-Hill.
SAPIN, C., FANTINO, B., NOWICKI, M.-L. & KIND, P. 2004. Usefulness of EQ-5D in assessing health status in primary care patients with major depressive disorder. *Health and Quality of Life Outcomes,* 2**,** 20.
SCHWANDER, T. & LEIMAR, O. 2011. Genes as leaders and followers in evolution. *Trends in Ecology & Evolution,* 26**,** 143-151.
SCHWARTZ, J. H. 1999. Homeobox genes, fossils, and the origin of species. *The Anatomical Record,* 257**,** 15-31.
SCULPHER, M., CLAXTON, K. & AKEHURST, R. 2005. It's just evaluation for decision making: recent developments in, and challenges for, cost-effectiveness research. *Health Policy and Economics. Opportunities and Challenges***,** 8-41.
SCULPHER, M. J., CLAXTON, K., DRUMMOND, M. & MCCABE, C. 2006. Whither trial-based economic evaluation for health care decision making? *Health economics,* 15**,** 677-687.
SEGAL, S. & POLLARD, A. 2004. Vaccines against bacterial meningitis. *British medical bulletin,* 72**,** 65-81.
SEN, A. 2008. The economics of happiness and capability. *Capabilities and happiness,* 27.
SERDULA, M., WILLIAMSON, D., KENDRICK, J., ANDA, R. & BYERS, T. 1991. Trends in alcohol consumption by pregnant women. *Journal of the American Medical Association,* 265**,** 876-879.
SEVERENS, J. L. & MILNE, R. J. 2004. Discounting health outcomes in economic evaluation: the ongoing debate. *Value in health,* 7**,** 397-401.
SHARP, D., CHEW-GRAHAM, C., TYLEE, A., LEWIS, G., HOWARD, L., ANDERSON, I., ABEL, K., TURNER, K., HOLLINGHURST, S. & TALLON, D. 2010. A pragmatic randomised controlled trial to compare antidepressants with a community-based psychosocial intervention for the treatment of women with postnatal depression: the RESPOND trial. *Health Technol Assess,* 14**,** 1-153.
SHARPE, D. 1997. Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical psychology review,* 17**,** 881-901.
SHLENS, J. 2014. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.
SMITH, P. K., COWIE, H. & BLADES, M. 2015. *Understanding children's development*, John Wiley & Sons.
SOARES, M. O., DUMVILLE, J. C., ASHBY, R. L., IGLESIAS, C. P., BOJKE, L., ADDERLEY, U., MCGINNIS, E., STUBBS, N., TORGERSON, D. J. & CLAXTON, K. 2013. Methods to assess cost-effectiveness and value of further research when data are sparse: negative-pressure wound therapy for severe pressure ulcers. *Medical Decision Making,* 33**,** 415-436.
SOARES, M. O., SHARPLES, L., MORTON, A., CLAXTON, K. & BOJKE, L. 2018. Experiences of Structured Elicitation for Model-Based Cost-Effectiveness Analyses. *Value in health*.
STEELE, F. 2008. Multilevel models for longitudinal data. *Journal of the Royal Statistical Society: series A (statistics in society),* 171**,** 5-19.
STEGER, M. F. 2006. An illustration of issues in factor extraction and identification of dimensionality in psychological assessment data. *Journal of personality Assessment,* 86**,** 263-272.

STEWART, D.E., ROBERTSON, E., DENNIS, C.L., GRACE, S.L. & WALLINGTON, T., 2003. Postpartum depression: Literature review of risk factors and interventions. *Toronto: University Health Network Women's Health Program for Toronto Public Health*.
STODDART, G. & EVANS, R. 2017. Producing health, consuming health care. *Why are some people healthy and others not?* : Routledge.
SULLIVAN, P. W., SLEJKO, J. F., SCULPHER, M. J. & GHUSHCHYAN, V. 2011. Catalogue of EQ-5D scores for the United Kingdom. *Medical Decision Making,* 31**,** 800-804.
TAMMEN, S. A., FRISO, S. & CHOI, S.-W. 2013. Epigenetics: the link between nature and nurture. *Molecular aspects of medicine,* 34**,** 753-764.
TANNER, J. M. 1990. *Foetus Into Man: Physical Growth from Conception to Maturity*, Harvard University Press.
TENNANT, R., HILLER, L., FISHWICK, R., PLATT, S., JOSEPH, S., WEICH, S., PARKINSON, J., SECKER, J. & STEWART-BROWN, S., 2007. The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health and Quality of life Outcomes*, *5*(1), p.63.
THIEFFRY, D. & SARKAR, S. 1998. Forty years under the central dogma. *Trends in biochemical sciences,* 23**,** 312-316.
TONER, R., SNAPE, C., ACTON, S. & BLENKIRON, P., 2010. Do general practitioners adhere to NICE guidelines for depression? Systematic Questionnaire Survey. *Primary Health Care Research & Development*, *11*(2), pp.123-131.
UNGAR, W. 2009. *Economic evaluation in child health*, Oxford University Press.
UOL (UNIVERSITY OF LONDON. INSTITUTE OF EDUCATION. CENTRE FOR LONGITUDINAL STUDIES), 2015a. Millennium Cohort Study: Linked Education Administrative Dataset (KS1), England: Secure Access. [data collection]. 2nd Edition. UK Data Service. SN:6862, http://doi.org/10.5255/UKDA-SN-6862-3.
UOL (UNIVERSITY OF LONDON. INSTITUTE OF EDUCATION. CENTRE FOR LONGITUDINAL STUDIES), 2015b. Millennium Cohort Study: Linked Education Administrative Dataset (KS2), England: Secure Access. [data collection]. UK Data Service. SN:7712, http://doi.org/10.5255/UKDA-SN-7712-1.
UOL (UNIVERSITY OF LONDON. INSTITUTE OF EDUCATION. CENTRE FOR LONGITUDINAL STUDIES), 2016. 1970 British Cohort Study: Waves 1-9, 1970-2012. 2nd Edition. UK Data Service. Retreived from https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7473.
UOL (UNIVERSITY OF LONDON. INSTITUTE OF EDUCATION. CENTRE FOR LONGITUDINAL STUDIES), 2017. Millennium Cohort Study: Waves 1-5, 2000-2011. 4th Edition. UK Data Service. Retreived from https://discover.ukdataservice.ac.uk/catalogue/?sn=4683&type=Data%20catalogue.
UK NCS (THE UK NATIONAL SCREENING COMMITTEE), 2011. *The UK NSC recommendation on Postnatal depression screening in pregnancy* [Online]. Available: https://legacyscreening.phe.org.uk/postnataldepression 2018].
UK NCS (THE UK NATIONAL SCREENING COMMITTEE), 2015. *Criteria for appraising the viability, effectiveness and appropriateness of a screening program* [Online]. Public Health England GOV.UK. Available: https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme 2018].
URDAHL, H., MANCA, A. & SCULPHER, M. J. 2006. Assessing generalisability in model-based economic evaluation studies. *Pharmacoeconomics,* 24**,** 1181-1197.
VAN DEN DRIES, L., JUFFER, F., VAN IJZENDOORN, M. H. & BAKERMANS-KRANENBURG, M. J. 2009. Fostering security? A meta-analysis of attachment in adopted children. *Children and youth services review,* 31**,** 410-421.
VANDEKERCKHOVE, J., MATZKE, D. & WAGENMAKERS, E.-J. 2015. Model Comparison and the Principle. *The Oxford handbook of computational and mathematical psychology***,** 300.

VLIEGEN, N., CASALIN, S. & LUYTEN, P. 2014. The course of postpartum depression: a review of longitudinal studies. *Harvard review of psychiatry,* 22**,** 1-22.
WADSWORTH, M., KUH, D., RICHARDS, M. & HARDY, R. 2005. Cohort profile: the 1946 national birth cohort (MRC National Survey of Health and Development). *International journal of epidemiology,* 35**,** 49-54.
WALKER, S. P., WACHS, T. D., GARDNER, J. M., LOZOFF, B., WASSERMAN, G. A., POLLITT, E., CARTER, J. A. & GROUP, I. C. D. S. 2007. Child development: risk factors for adverse outcomes in developing countries. *The lancet,* 369**,** 145-157.
WALSHE, K., COLEMAN, A., MCDONALD, R., LORNE, C. & MUNFORD, L., 2016. Health and social care devolution: the Greater Manchester experiment. *Bmj*, *352*, p.i1495.
WANG, Y. C., MCPHERSON, K., MARSH, T., GORTMAKER, S. L. & BROWN, M. 2011. Health and economic burden of the projected obesity trends in the USA and the UK. *The Lancet,* 378**,** 815-825.
WEBB, P., BAIN, C. & PAGE, A. 2016. *Essential epidemiology: an introduction for students and health professionals*, Cambridge University Press.
WEBBER, A. L., WOOD, J. M., GOLE, G. A. & BROWN, B. 2008. The effect of amblyopia on fine motor skills in children. *Investigative ophthalmology & visual science,* 49**,** 594-603.
WEIJERMAN, M. E. & DE WINTER, J. P. 2010. Clinical practice. *European journal of pediatrics,* 169**,** 1445-1452.
WEINFIELD, N. S., SROUFE, L. A. & EGELAND, B. 2000. Attachment from infancy to early adulthood in a high-risk sample: Continuity, discontinuity, and their correlates. *Child development,* 71**,** 695-702.
WHITE, I. R., ROYSTON, P. & WOOD, A. M. 2011. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine,* 30**,** 377-399.
WHITEHEAD, S. J. & ALI, S. 2010. Health outcomes in economic evaluation: the QALY and utilities. *British medical bulletin,* 96**,** 5-21.
WHO (THE WORLD HEALTH ORGANIZATION). 2018. Health Technology Assessment.
WILLIAMS, R. 2012. Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal,* 12**,** 308.
WINDSOR, R. A., WOODBY, L. L., MILLER, T. M., HARDIN, J. M., CRAWFORD, M. A. & DICLEMENTE, C. C. 2000. Effectiveness of Agency for Health Care Policy and Research clinical practice guideline and patient education methods for pregnant smokers in Medicaid maternity care. *American Journal of Obstetrics & Gynecology,* 182**,** 68-75.
WOODS, B., REVILL, P., SCULPHER, M. & CLAXTON, K. 2016. Country-level cost-effectiveness thresholds: initial estimates and the need for further research. *Value in Health,* 19**,** 929-935.
WRIGHT, B., BARRY, M., HUGHES, E., TRÉPEL, D., ALI, S., ALLGAR, V., COTTRILL, L., DUFFY, S., FELL, J. & GLANVILLE, J. 2015. Clinical effectiveness and cost-effectiveness of parenting interventions for children with severe attachment problems: a systematic review and meta-analysis.
WSIPP (WASHINGTON STATE INSTITUTE FOR PUBLIC POLICY), 2017. *Benefit-Cost Technical Documentation. Washington State Institute for Public Policy Benefit-Cost Model* [Online]. Available: http://www.wsipp.wa.gov/BenefitCost [Accessed 02/08/2018.
YOUNG, R. & JOHNSON, D. R. 2015. Handling missing values in longitudinal panel data with multiple imputation. *Journal of Marriage and Family,* 77**,** 277-294.

References: Chapter 4 Scoping Review Included Studies

ADAIR, L. S., FALL, C. H., OSMOND, C., STEIN, A. D., MARTORELL, R., RAMIREZ-ZEA, M., SACHDEV, H. S., DAHLY, D. L., BAS, I., NORRIS, S. A., MICKLESFIELD, L., HALLAL, P., VICTORA, C. G. & GROUP, C. 2013. Associations of linear growth and relative weight gain during early life with adult health and human capital in countries of low and middle income: findings from five birth cohort studies. Lancet, 382, 525-34.

BOYLE, M. H., GEORGIADES, K., RACINE, Y. & MUSTARD, C. 2007. Neighborhood and family influences on educational attainment: Results from the Ontario Child Health Study Follow-Up 2001. Child Development, 78, 168-189.

CHANDOLA, T., CLARKE, P., MORRIS, J. N. & BLANE, D. 2006. Pathways between Education and Health: A Causal Modelling Approach. Journal of the Royal Statistical Society: Series A (Statistics in Society), 169, 337-59.

CHARTIER, M. J., WALKER, J. R. & NAIMARK, B. 2010. Separate and cumulative effects of adverse childhood experiences in predicting adult health and health care utilization. Child Abuse & Neglect, 34, 454-464.

CREEL, M. & FARELL, M. 2006.The black-white test score gap widens with age? UFAE and IAE Working Papers 670.06, Unitat de Fonaments de l'Anàlisi Econòmica (UAB) and Institut d'Anàlisi Econòmica (CSIC).

CUNHA, F. & HECKMAN, J., 2008. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. Journal of human resources, 43(4), pp.738-782.

CUNHA, F. & HECKMAN, J. 2014. Estimating the technology of cognitive and noncognitive skill formation: The linear case. Molenaar, Peter C M [Ed]; Lerner, Richard M [Ed]; Newell, Karl M [Ed] (2014) Handbook of developmental systems theory and methodology (pp 221-269) ix, 517 pp New York, NY, US: Guilford Press; US, 221-269.

DANIELS, M. C. & ADAIR, L. S. 2004. Growth in young Filipino children predicts schooling trajectories through high school. Journal of Nutrition, 134, 1439-1446.

DISHION, T. J., VERONNEAU, M. H. & MYERS, M. W. 2010. Cascading peer dynamics underlying the progression from problem behavior to violence in early to late adolescence. Development and psychopathology, 22, 603-619.

DUBOW, E. F., BOXER, P. & ROWELL, L. 2009. Long-term Effects of Parents' Education on Children's Educational and Occupational Success. Merrill-Palmer Quarterly, 55, 224-249.

ENGLE, P. L., FERNALD, L. C., ALDERMAN, H., BEHRMAN, J., O'GARA, C., YOUSAFZAI, A., DE MELLO, M. C., HIDROBO, M., ULKUER, N., ERTEM, I., ILTUS, S. & GLOBAL CHILD DEVELOPMENT STEERING, G. 2011. Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. Lancet, 378, 1339-53.

FERGUSSON, D., SWAIN-CAMPBELL, N. & HORWOOD, J. 2004. How does childhood economic disadvantage lead to crime? Journal of Child Psychology & Psychiatry & Allied Disciplines, 45, 956-66.

FERRER, E. & MCARDLE, J. J. 2004. An Experimental Analysis of Dynamic Hypotheses about Cognitive Abilities and Achievement from Childhood to Early Adulthood. Developmental Psychology, 40, 935-952.

FRIEDMAN, H. S. 2014. Integrating prospective longitudinal data: modeling personality and health in the Terman Life Cycle and Hawaii Longitudinal Studies. Developmental psychology, 50, 1390-1406.

FRIJTERS, P., HATTON, T. J., MARTIN, R. M. & SHIELDS, M. A. 2010. Childhood economic conditions and length of life: Evidence from the UK Boyd Orr cohort, 1937-2005. Journal of Health Economics, 29, 39-47.

HAGGER-JOHNSON, G., BATTY, G., DEARY, I. J. & VON STUMM, S. 2011. Childhood socioeconomic status and adult health: Comparing formative and reflective models in the Aberdeen Children of the 1950s Study (prospective cohort study). Journal of Epidemiology and Community Health, 65, 1024-1029.
HAGGER-JOHNSON, G., MOTTUS, R., CRAIG, L. C., STARR, J. M. & DEARY, I. J. 2012. Pathways from childhood intelligence and socioeconomic status to late-life cardiovascular disease risk. Health Psychology, 31, 403-412.
HAMPSON, S. E., EDMONDS, G. W., GOLDBERG, L. R., DUBANOSKI, J. P. & HILLIER, T. A. 2015. A life-span behavioral mechanism relating childhood conscientiousness to adult clinical health. Health Psychology, 34, 887-895.
HATCH, S. L., HARVEY, S. B. & MAUGHAN, B. 2010. A developmental-contextual approach to understanding mental health and well-being in early adulthood. Social Science & Medicine, 70, 261-8.
HEALEY, A., KNAPP, M. & FARRINGTON, D. P. 2004. Adult Labour Market Implications of Antisocial Behaviour in Childhood and Adolescence: Findings from a UK Longitudinal Study. Applied Economics, 36, 93-105.
HERRENKOHL, T. I., KOSTERMAN, R., MASON, W. A., HAWKINS, J. D., MCCARTY, C. A. & MCCAULEY, E. 2010. Effects of childhood conduct problems and family adversity on health, health behaviors, and service use in early adulthood: tests of developmental pathways involving adolescent risk taking and depression. Development and psychopathology, 22, 655-665.
HERTZMAN, C., POWER, C., MATTHEWS, S. & MANOR, O. 2001. Using an interactive framework of society and lifecourse to explain self-rated health in early adulthood. Social Science and Medicine, 53, 1575-1585.
HUANG, C., SOLDO, B. J. & ELO, I. T. 2011. Do early-life conditions predict functional health status in adulthood? The case of Mexico. Social Science and Medicine, 72, 100-107.
JIMERSON, S. R., EGELAND, B., SROUFE, L. & CARLSON, B. 2000. A prospective longitudinal study of high school dropouts: Examining multiple predictors across development. Journal of School Psychology, 38, 525-549.
JOHNSON, W., MCGUE, M. & IACONO, W. G. 2006. Genetic and environmental influences on academic achievement trajectories during adolescence. Developmental Psychology, 42, 514-532.
KUH, D., RICHARDS, M., HARDY, R., BUTTERWORTH, S. & WADSWORTH, M. E. J. 2004. Childhood cognitive ability and deaths up until middle age: A post-war birth cohort study. International Journal of Epidemiology, 33, 408-413.
KUH, D., SHAH, I., RICHARDS, M., MISHRA, G., WADSWORTH, M. & HARDY, R. 2009. Do childhood cognitive ability or smoking behaviour explain the influence of lifetime socio-economic conditions on premature adult mortality in a British post war birth cohort? Social Science & Medicine, 68, 1565-73.
LACOURSE, E., COTE, S., NAGIN, D. S., VITARO, F., BRENDGEN, M. & TREMBLAY, R. E. 2002. A longitudinal-experimental approach to testing theories of antisocial behavior development. Development and Psychopathology, 14, 909-924.
LACOURSE, E., NAGIN, D. S., VITARO, F., COTE, S., ARSENEAULT, L. & TREMBLAY, R. E. 2006. Prediction of Early-Onset Deviant Peer Group Affiliation: A 12-Year Longitudinal Study. Archives of General Psychiatry, 63, 562-568.
LAGER, A. C. J., MODIN, B. E., DE STAVOLA, B. L. & VAGERO, D. H. 2012. Social origin, schooling and individual change in intelligence during childhood influence long-term mortality: A 68-year follow-up study. International Journal of Epidemiology, 41, 398-404.
LAYARD, R., CLARK, A. E., CORNAGLIA, F., POWDTHAVEE, N. & VERNOIT, J. 2014. What Predicts a Successful Life? A Life-Course Model of Well-Being. Economic Journal, 124, F720-38.

LEE, T., WICKRAMA, K. & SIMONS, L. 2013. Chronic Family Economic Hardship, Family Processes and Progression of Mental and Physical Health Symptoms in Adolescence. Journal of Youth & Adolescence, 42, 821-836.
LINDEBOOM, M., LLENA-NOZAL, A. & VAN DER KLAAUW, B. 2006. Disability and Work: The Role of Health Shocks and Childhood Circumstances. CEPR Discussion Paper No. 5685. Available at SSRN: https://ssrn.com/abstract=921840
MARTENS, P. J., CHATEAU, D. G., BURLAND, E. M., FINLAYSON, G. S., SMITH, M. J., TAYLOR, C. R., BROWNELL, M. D., NICKEL, N. C., KATZ, A. & BOLTON, J. M. 2014. The effect of neighborhood socioeconomic status on education and health outcomes for children living in social housing. American Journal of Public Health, 104, 2103-2113.
MASON, P. L. 2007. Intergenerational Mobility and Interracial Inequality: The Return to Family Values. Industrial Relations: A Journal of Economy & Society, 46, 51-80.
MASON, W. A., HITCH, J. E., KOSTERMAN, R., MCCARTY, C. A., HERRENKOHL, T. I. & HAWKINS, J. D. 2010. Growth in adolescent delinquency and alcohol use in relation to young adult crime, alcohol use disorders, and risky sex: a comparison of youth from low- versus middle-income backgrounds W. Alex Mason et al. Adolescent delinquency and alcohol use. Journal of Child Psychology & Psychiatry, 51, 1377-1385.
MOFFITT, T. E., ARSENEAULT, L., BELSKY, D., DICKSON, N., HANCOX, R. J., HARRINGTON, H., HOUTS, R., POULTON, R., ROBERTS, B. W., ROSS, S., SEARS, M. R., THOMSON, W. & CASPI, A. 2011. A gradient of childhood self-control predicts health, wealth, and public safety. PNAS Proceedings of the National Academy of Sciences of the United States of America, 108, 2693-2698.
MOODY-AYERS, S., LINDQUIST, K., SEN, S. & COVINSKY, K. E. 2007. Childhood social and economic well-being and health in older age. American Journal of Epidemiology, 166, 1059-1067.
MUENNIG, P. 2009. The social costs of childhood lead exposure in the post-lead regulation era. Archives of Pediatrics and Adolescent Medicine, 163, 844-849.
NANDI, A., GLYMOUR, M. M., KAWACHI, I. & VANDERWEELE, T. J. 2012. Using marginal structural models to estimate the direct effect of adverse childhood social conditions on onset of heart disease, diabetes, and stroke. Epidemiology, 23, 223-232.
NIKULINA, V., WIDOM, C. S. & CZAJA, S. 2011. The Role of Childhood Neglect and Childhood Poverty in Predicting Mental Health, Academic Achievement and Crime in Adulthood. American Journal of Community Psychology, 48, 309-321.
PETRAS, H., SCHAEFFER, C. M., IALONGO, N., HUBBARD, S., MUTHEN, B., LAMBERT, S. F., PODUSKA, J. & KELLAM, S. 2004. When the course of aggressive behavior in childhood does not predict antisocial outcomes in adolescence and young adulthood: An examination of potential explanatory variables. Development and Psychopathology, 16, 919-941.
REYNOLDS, A. J. & OU, S. R. 2011. Paths of Effects From Preschool to Adult Well-Being: A Confirmatory Analysis of the Child-Parent Center Program. Child Development, 82, 555-582.
RISI, S., GERHARDSTEIN, R. & KISTNER, J. 2003. Children's Classroom Peer Relationships and Subsequent Educational Outcomes. Journal of Clinical Child and Adolescent Psychology, 32, 351-361.
ROESER, R. W. & PECK, S. C. 2003. Patterns and pathways of educational achievement across adolescence: a holistic-developmental perspective. New directions for child and adolescent development, 39-62.
ROSA DIAS, P. 2009. Inequality of opportunity in health: evidence from a UK cohort study. Health Economics, 18, 1057-74.
SAVAGE, J. & VILA, B. 2002. Changes in child welfare and subsequent crime rate trends: A cross-national test of the lagged nurturance hypothesis. Journal of Applied Developmental Psychology, 23, 51-82.

SCHOON, I., SACKER, A. & BARTLEY, M. 2003. Socio-economic adversity and psychosocial adjustment: A developmental-contextual perspective. Social Science and Medicine, 57, 1001-1015.
SHARIEFF, W., ZLOTKIN, S. H., UNGAR, W. J., FELDMAN, B., KRAHN, M. D. & TOMLINSON, G. 2008. Economics of preventing premature mortality and impaired cognitive development in children through home-fortification: a health policy perspective. International Journal of Technology Assessment in Health Care, 24, 303-11.
SHEN, K. & ZENG, Y. 2014. Direct and indirect effects of childhood conditions on survival and health among male and female elderly in China. Social Science & Medicine, 119, 207-214.
SLOPEN, N., NON, A., WILLIAMS, D. R., ROBERTS, A. L. & ALBERT, M. A. 2014. Childhood adversity, adult neighborhood context, and cumulative biological risk for chronic diseases in adulthood. Psychosomatic Medicine, 76, 481-489.
TE VELDE, S. J., LENNERT VEERMAN, J., TAK, N. I., BOSMANS, J. E., KLEPP, K. I. & BRUG, J. 2011. Modeling the long term health outcomes and cost-effectiveness of two interventions promoting fruit and vegetable intake among schoolchildren. Economics & Human Biology, 9, 14-22.
TRAN, B. X., OHINMAA, A., KUHLE, S., JOHNSON, J. A. & VEUGELERS, P. J. 2014. Life course impact of school-based promotion of healthy eating and active living to prevent childhood obesity. PLoS ONE, 9.
TUBEUF, S., JUSOT, F. & BRICARD, D. 2012. Mediating role of education and lifestyles in the relationship between early-life conditions and health: Evidence from the 1958 British Cohort. Health Economics, 21, 129-150.
VARTANIAN, T. P. & BUCK, P. W. 2005. Childhood and Adolescent Neighborhood Effects on Adult Income: Using Siblings to Examine Differences in Ordinary Least Squares and Fixed-Effect Models. Social Service Review, 79, 60-94.
WODTKE, G. T., HARDING, D. J. & ELWERT, F. 2016. Neighborhood Effect Heterogeneity by Family Income and Developmental Period. AJS - American Journal of Sociology, 121, 1168-222.
WORLD BANK. 2003. Caribbean youth development: Issues and policy directions, World Bank Country Studies. Washington, D.C., Author.