Producing effective and achievable safety strategies from adverse event investigations in healthcare.

William Patrick Lea

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Healthcare, Faculty of Medicine and Health

July 2025

Intellectual property and publications statement

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 3 has been published[¥]: Lea W, Lawton R, Vincent C, O'Hara, J. Exploring the "Black Box" of Recommendation Generation in Local Health Care Incident Investigations: A Scoping Review. Journal of Patient Safety 19(8):p 553-563, December 2023.

The candidate (WL) developed the study design, search strategy, and data extraction strategy with support from JOH, RL and CV. WL undertook the searches. Titles and abstracts were reviewed by WL, and a random sample of 5% were screened by both JOH and RL independently. WL reviewed an undertook the study quality/risk of bias assessments with support from JOH and RL. Analysis was undertaken by WL with support from JOH, RL and CV. WL drafted the manuscript, and edited with comments from peer review from two journals, with input and guidance from JOH, RL and CV at each stage.

Chapter 4 has been published[¥]: Lea W, Budworth L, O'Hara J, Vincent C, Lawton R. Investigators are human too: outcome bias and perceptions of individual culpability in patient safety incident investigations *BMJ Quality & Safety* Published Online First: 10 February 2025.

WL developed the study design with support from JOH, RL and CV. WL led every step of the research with support JOH, RL and CV as well an external advisor (Luke Budworth, LB). WL drafted the manuscript, and edited with comments from journal peer review, with input and guidance from JOH, RL, LB and CV at each stage. LB provided guidance on the statistical analysis plan and data analysis.

Chapter 5 has been drafted for submission to a suitable journal

WL developed the study design with support from JOH, RL and CV as well as an external expert steering group (Claire Cox(CC), Sarah Seddon(SS) and Paul Bowie(PB)). WL led every step of the research with support JOH, RL, CV, CC, SS and PB. WL drafted the manuscript, and edited with input and guidance from JOH, RL, and CV.

Copyright statement

Funding statement

This research was funded by York & Scarborough Teaching Hospitals Charity and Trust. The views expressed are those of the author and not necessarily those of the Charity or Trust. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

First and foremost, I would like to thank my fantastic supervisors Jane, Rebecca, and Charles. Not only are they lovely people but inspirational academics. They have given me their time and wisdom as well as encouragement, supportive critique and care.

I would like to thank Donald Richardson, Diane Palmer, Jonathan Thow, and Vijay Jayagopal who have all had significant roles in shaping my interest in patient safety as well as encouraging and supporting me throughout my PhD.

The York Hospital Charity and York & Scarborough Teaching Hospital NHSFT have provided me with generous financial support to undertake this PhD.

I have connected with many wonderful people during my PhD. I would like to thank these people and others for their time and valuable feedback, suggestions and support.

While not an exhaustive list some of the people I have connected with include: Alan Card (Assistant Professor at the San Diego State University School of Public Health), Claire Cox (Experienced Nurse, Patient Safety Manager, Associate Director of Patient Safety Learning, UK), Helen Hughes (Chief Executive of Patient Safety Learning, UK), Peter Hibbert (Australian Institute of Health Innovation, Macquarie University), Farhad Peerally (Associate Professor/Honorary Consultant in Gastroenterology, UK), Carla Maclean (Kwantlen Polytechnic University, Canada), Sarah Seddon (Specialist Pharmacist, and Harmed Patient Alliance, UK), Paul Bowie (Programme Director (Safety & Improvement) Scotland), Livi Joseph (Senior Equity Research Fellow and PhD student, UK), Rosemary Lim (Associate Professor, University of Reading, UK), Alex Carter (Associate Professor, London School of Economics and Political Science, UK), Matthew Wain (Deputy Director of Investigations, HSSIB, UK), Andrew Murphy-Pittock (Director, HSSIB, UK), Richard McMaster (Senior Investigation Science Educator, HSSIB, UK). Regular or singular discussions with these people, as well as others, throughout my PhD has provided continuous feedback, intellectual challenge, and ‘reality check’.

Abstract

It is estimated that one in ten patients will experience an 'adverse event'; that is something going wrong in the way in which care is delivered. These events are often investigated and recommendations made, with the intention of preventing recurrence or improving safety. There are, however, increasing concerns that these recommendations and preceding investigations are not contributing to improved safety, and potentially contributing to safety clutter. The aim of this PhD was to explore how the generation of recommendations might be improved. Three studies were undertaken: 1) a scoping review; 2) an experimental scenario study; and 3) a modified Delphi study. **Study 1** (scoping review) highlighted that recommendations tended to focus on individuals' behaviour rather than latent system deficiencies, with a lack of agreement about how recommendations should be judged for effectiveness. These two findings led onto the subsequent studies.

Firstly, given the scoping review findings that investigation recommendations seem to 'blame' the actions of individuals, and focus improvement efforts on changing their behaviour, the possibility that cognitive biases of those involved in investigations may play a role in this tendency was explored. **Study 2** was an experimental scenario study, designed to examine the impact of outcome bias on judgements of staff responsibility, incident avoidability, importance of investigating and recommendation selection. Outcome bias occurs when the ultimate outcome of a past event is given excessive weight, in comparison to other information, when judging the preceding actions or decisions. The results of this study indicated that outcome bias had significant impact of judgement and responses when investigating incidents, with higher ratings of staff responsibility, importance of investigating and higher likelihood of punitive recommendations when patients came to greater harm. While expertise in safety reduced this impact it did not entirely eliminate it.

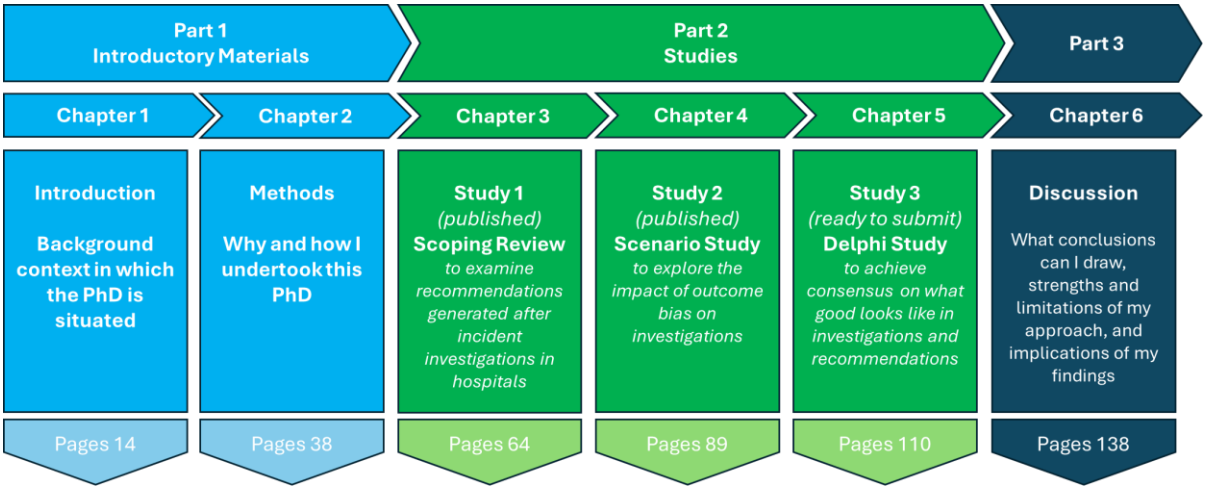
Secondly, **Study 1** findings suggested difficulties in judging recommendations' quality or effectiveness, and that there was no consistent approach in the literature. Before attempting to improve recommendations, it is first necessary to define what a 'good recommendation' is. **Study 3** was a modified Delphi study that aimed to achieve

consensus on what 'good' looks like in investigation and recommendation generation. As recommendations are closely linked to the findings and activities of the investigation, it was decided to attempt to gain consensus on criteria to judge both the quality of an investigation and recommendations. Ninety-two evidenced-based criteria were drafted with the help of an expert steering group. Following three rounds of the Delphi process, consensus was achieved for 92 criteria, which were then ranked by their ratings and level of expert agreement. Further work is needed to understand how these criteria could be used to judge and improve the quality of investigations and recommendations.

Taken together, this evidence suggests that the generation of recommendations is a complex, and that the current evidence does not sufficiently describe how this important work is achieved in everyday healthcare practice. What is evident is that despite the increasing awareness of systems factors in the incidence of adverse events, outcome bias is a significant influence on the generation of recommendations and the assignment of responsibility. With the potentially far-reaching impact of cognitive biases on investigations and recommendation generation, further work is needed to examine the impact as well as mitigation strategies. The generation of recommendations is further complicated by the lack of guidance about what best practice might be in the investigation and recommendation generation process. Indeed, cognitive bias identification and mitigation is but one of the criteria formulated from the modified Delphi study. These criteria will be useful in measuring the effectiveness of investigations and recommendations within future research but also for front-line teams in patient safety. However, further research will be needed to understand how these criteria can be operationalised for systematic application by healthcare staff, to improve their processes for learning from patient safety events.

Page

Contents



Tables

Page

Figures

Page

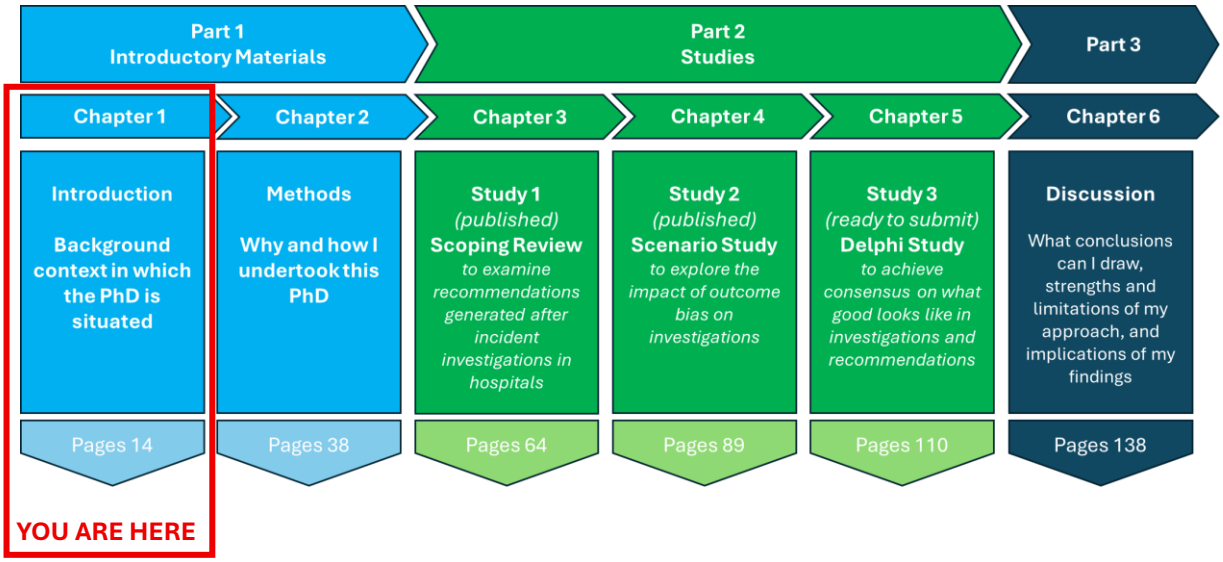
Ethical Approvals

Study 2, Chapter 4 was reviewed and approved by the School of Healthcare Research Ethics Committee (SHREC), University of Leeds (HREC 21-013).

Study 3, Chapter 5 was reviewed and approved by the School of Healthcare Research Ethics Committee (SHREC), University of Leeds (HREC 23-002).

Part 1 - Introductory Materials

Part 1 of this thesis locates the work within the reviewed literature and practical context of hospital patient safety investigations and recommendations (Chapter 1). Chapter 2 presents the approach to research and how the research questions were devised and developed, aims and objectives, and the methods that were used to try and answer these questions. As illustrated in the diagram below, part 2 contains the PhD studies, and is followed by part 3 which contains a critical discussion of the published studies within the wider academic and practical context.



Chapter 1 Introduction

The extent and impact of unsafe care was highlighted in seminal reports published in the USA and UK, which called for the establishment of systems to learn and improve.[1,2] Since then, the majority of efforts to improve patient safety in hospitals, have revolved around identifying occurrences of unsafe care (incident reporting), followed by investigation of the potential causes, and then production of recommendations aimed at reducing risk and future harm.[3] The aspiration was that healthcare could adopt principles of learning from incidents from the aviation industry in order to achieve the kinds of improvements seen in aviation safety.[3] While there have been some notable examples of improvements as a result of reporting systems and investigations,[4-8] the overall impact has fallen short of expectations.[9-13] There are increasing concerns that

investigations and subsequent recommendations, are themselves, not contributing to improved safety.[9-12,14]

In particular, the recommendations arising from incident investigations, have come under increasing academic scrutiny.[10-12,15] This interest has occurred in parallel with the establishment of national-level independent investigatory bodies (e.g., HSSIB in the UK, Norwegian Healthcare Investigation Board in Norway),[16,17] and in the UK, an ever increasing number of public inquiries, such as the Morecambe Bay Investigation, and the ever expanding set of associated recommendations.[18] Therefore, exploring the act of recommendation generation is of increasing relevance as the number of recommendations across both local and national level investigation activity grows exponentially.

The generation of recommendations within hospitals is the focus of this thesis. Chapter 1 will provide background and context to patient safety, the history of incident reporting, investigations, and finally, the generation of recommendations.

1.1 Healthcare's awareness of the extent of harm from unsafe care

"First, do no harm" is a fundamental principle of healthcare, yet aside from a few sporadic publications,[19-23] until the latter part of the 20th Century there had been a distinct lack of academic or practical attention to patient harm. There was little, in the published literature, about the nature of iatrogenic harm until the 1990s.[2]

Following a number of studies in 1991 reporting significant and potentially preventable iatrogenic patient harm in US hospitals,[24-26] Lucian Leape, a paediatric surgeon in the US, published an important paper, "Error in Medicine", in which he highlighted the potential significance of patient harm.[27] Leape discussed the concepts of preventability and human error as well as the important part that psychology and human factors might play in reducing harm.[27] Within a few years the United States of America's Institute of Medicine (IoM) published "To Err is Human: Building a Safer Health System", reporting that between 44,000 and 98,000 died each year in American hospitals as a result of

medical errors.[2] Shortly after the IoMs' report The UK government published "An organisation with a memory" which reported, among many other statistics, that 10% of patients admitted to hospitals experienced harm.[1] These reports received a great deal of attention, both from the public and global healthcare community, and are often credited with spawning the 'patient safety movement.[28] In the 10 years following their publication, over 5000 articles were published on patient safety and medical errors.[29]

The concepts of unsafe care and harm are central to this PhD and a number of terms need to be defined. When **defining unsafe care**, Emanuel and colleagues definition of patient safety or safe care is useful:

"Patient safety is a discipline in the health care sector that applies safety science methods toward the goal of achieving a trustworthy system of health care delivery. Patient safety is also an attribute of health care systems; it minimizes the incidence and impact of, and maximizes recovery from, adverse events."[30: p.6]

Unsafe care can therefore be defined as care that does not meet this definition, i.e. does not minimize the incidence or impact of, or maximise recovery from, adverse events. It is in fact likely that unsafe care may increase the incidence and impact of adverse events. It is important here to point out that unsafe care does not necessarily result in harm. **Patient safety incidents (PSIs)** have been described as events that may or may not result in harm. For example, a study by Baker et al found that 35% of PSIs examined resulted in no harm,[31] and Sari and colleagues found 23% of admissions experienced PSIs, while 11% of admissions experienced harm as a result of PSIs.[32]

Adverse events (AEs) are defined as unintended injury to a patient that results in harm, disability, death, or prolonged hospital stay and result from healthcare management, or the way in which healthcare is delivered, rather than any underlying disease process.[25,31,33-36]

The impact of harm will be explored in the following section.

1.2 The impact of unsafe care, and importance of patient safety

The global impact of unsafe care is considerable. The World Health Organization in its 2024 report, reaffirmed the estimate that more than one in ten patients experience harm in healthcare settings.[36-38] Unsafe care results in significant morbidity and mortality with an estimated 3 million deaths per year, globally.[36,38]

Harm can be classified in a number of different ways, but commonly involves considering the severity of resulting physical symptoms or loss of function, duration of these, and the interventions required to treat or manage these symptoms or loss of function.[39] An important but often neglected consideration is the psychological harm that patients, and their relatives, may experience as a result of physical harm.[39-41]

Harm is not confined to the incident itself; the 'harm fallout' can be significant and far reaching. A growing body of evidence demonstrates how patients, their relatives, as well as healthcare staff may experience long-term physical and psychological harm following incidents and the subsequent investigations.[42-48] Indeed, attempts to learn from incidents and improve safety have been shown to, in some cases, compound the harm experienced by those involved, including both staff and patients.[40,49]

Beyond the physical and psychological harm to those involved in incidents, patient harm causes significant economic impact on both health systems and wider society. Most studies have been conducted within high-income countries, where estimates suggest 12.6% of total health expenditure goes towards managing the consequences of patient harm, which is equivalent to 1.4% of their combined gross domestic product.[36,50,51] The direct costs of patient harm are significant but may be smaller than the indirect costs of patient harm, such as the loss of peoples productivity, labour participation and income loss. While estimates should be cautiously interpreted, the downstream impact of patient harm is likely to be substantial at a societal level, and a growing body of evidence suggests inequality in its prevalence and impact.[38,52-57]

A greater awareness of the scale and impact of unsafe care, since the late 1990s, prompted a need to act. The following section presents one the main activities employed in an attempt to improve safety and reduce risk.

1.3 A short history of incident reporting systems

The seminal reports that launched the patient safety field (discussed in section 1.1)[1,2] called for organisations to learn from incidents, and to establish incident reporting systems(IRSs).[3,58] Within ten years, IRS were planned or in place in Australia, Austria, Belgium, Czech Republic, Denmark, Ireland, France, Netherlands, Norway, Scotland, Spain, Sweden, Switzerland, the United States of America(USA), and the United Kingdom (England and Wales).[58-64]

Doupi reported three different types of national patient safety IRSs:[58]

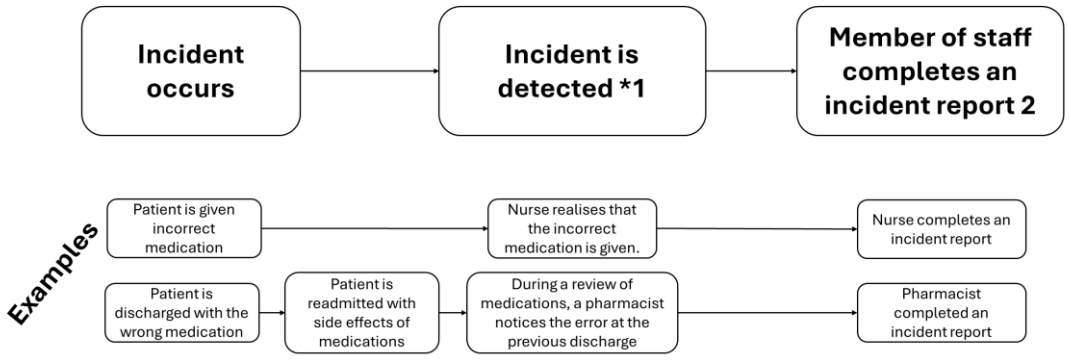
1) Those that, often by law, require the reporting of 'sentinel events', which the USAs Joint Commission describe as patient safety events that result in death, severe or permanent harm, or those events which, irrespective of harm, are considered important to investigate for the potential for learning that might reduce future harm.[58,65]
2) Those that focus on particular clinical domains, such as medication related incidents
3) Those that are healthcare system wide or national and include incidents with different levels of harm, no harm, or near misses (those incidents that could have resulted in harm but were detected and corrected before reaching a patient).

An example of the third type of IRS has existed in England since 2004. The UK National Patient Safety Agency (NPSA), established in 2001, launched the National Reporting and Learning System (NRLS) in 2004 - the first of its kind - to collect incident report data from across England and Wales. The purpose of collecting incident data was multi-factorial, partly as a requirement for local hospital learning, learning at a national level about the nature of patient safety incidents, and then finally the use of this information as the basis for action at a national level.[58,66] The NPSA ceased to exist in 2012, and some of its' activities have subsequently been taken over by NHS England.

National, regional or hospital policies will usually state that once an incident is identified, for instance by a member of staff, the first priority is reducing harm and risk to the patient involved.[58] As illustrated in figure 1.1 below, reporting of incidents is

largely completed by frontline healthcare staff who detect incidents during direct clinical activities or retrospective review (such as during hospital mortality review processes).[58] While early IRS may have used paper-based forms for recording incidents, hospitals increasingly utilise information technology with programs such as Datix – a system that allows staff to submit incident reports online.[58,67,68]

Figure 1.1 Detection and reporting of incidents.



*1 *There are a number of ways in which episodes of unsafe care may be detected: direct observation in clinical practice, by staff, team discussion, coroner's reports, clinical review meetings, death/mortality review processes/meetings, staff meeting discussions, complaints, patient, carer and family input (for example, questions, concerns), audits / monitoring variation in clinical practice, quality improvement work, chart reviews, whistleblowing, external investigations.*[69,70]

There are a number of issues that have been identified with incident reporting which bear relevance to this thesis. The information collected within incident reporting forms may vary between organisations; some favour brevity and ease of completion, while others require far more detail.[3,58] In fact the complexity of the incident reporting form and time to complete have been identified as important barriers to staff completing incident reports.[71] Macrae argues that initial reports may often be inaccurate or 'wrong', but might serve to trigger an investigation; and rather than working to improve the 'quality' or depth of incident reports, efforts should go into improving the quality of the investigations that follow.[3] This serves as one reason for the focus of this PhD on recommendations following investigations.

Not all episodes of unsafe care are detected, and a number of studies have demonstrated that IRSs only detect a minority of incidents (5-10%) that actually occur, potentially even less when incidents result in harm.[32,63,72-79]

To illustrate the rate at which incidents occur and are reported, data from the English NHS (Table 1.1) were used. Between April 2021 and March 2022 2,345,815 incidents were reported by healthcare staff across organisation in England (1,656,070 recorded as resulting in no harm, 608,959 low harm, 68,111 moderate harm, 6,872 severe harm, and 5,803 resulting in death).[80] Therefore if, as estimated, 5-10% of incidents that occur are reported, potentially 23,458,150 to 46,916,300 incidents actually occurred.

Table 1.1 Reported incidents and estimated total incidents (reported + unreported) in the English NHS between April 2021 and March 2022

	Incidents
Reported as no harm	1,656,070
Reported as low harm	608,959
Reported as moderate harm	68,111
Reported as severe harm	6,872
Reported as death	5,803
total incidents reported	2,345,815
total incidents that occurred if 10% are reported(Hibbert 2023)	23,458,150
total incidents that occurred if 5% are reported(Yu 2016)	46,916,300

A number of factors that encourage/support or hinder reporting have been identified and reported in the literature (Table 1.2). Despite the plethora of barriers and, as illustrated above, disconnect between incidents and reporting, the number of actual incidents reported has increased year on year. In England for example, the number of reported incidents has increased from under 100,000 in a 3-month period in 2003 to over 600,000 in a 3-month period in 2022.[80]

The administration of reporting systems is labour intensive and costly,[81-83] and this is before considering the costs and resource usage relating to subsequent investigations and implementation of recommendations. While incident reporting is likely to have contributed to improvements in safety,[6,7] there remains confusion about the purpose and practice of reporting.[3,84] The volume of reporting is a complex issue. While

organisations with higher reporting rates have been suggested to have a 'better safety culture',[77] rates and trends of reporting, often used as organisational safety data, are not considered a reliable indicator of safety or improvement.[3] Examining rates of particular types or locations of incidents may give a skewed view of safety within an organisation,[85] resulting in inappropriate focus of investigations or improvement efforts. Fluctuations in reporting rates, whether overall or for specific incident types, may reflect a change in actual incident rates, or simply changes in a range of other factors,[3] some of which are presented in Table 1.2. Another issue with incident reports is their content or 'quality'.[3] The contents of incident reports may be unreliable, biased or inaccurate.[3,85,86] Cooper and colleagues found that approximately half of incident reports contained judgements of blame towards individuals,[39] possibly neglecting the wider system issues that may have contributed to an incident.[87] Staff may submit mundane or trivial reports,[88] complaining about other staff or elements of the workplace they dislike.[89] These issues of 'data quality' may detract from the potential benefits of IRS.

While further work is needed to understand and define the purpose and processes around reporting in healthcare, Macrae argues that attention should be focused on what follows, namely "*the practical work of investigators is organised around finding and addressing gaps – gaps in how risks are represented and understood, and gaps in how safety is organised and improved.*".[90:p.193] It is this 'improvement' starting with recommendation generation, after investigations, that is the focus of this PhD.

Once an incident report has been completed the next step is for an organisation to decide how to respond to this, which I will discuss in the next section.

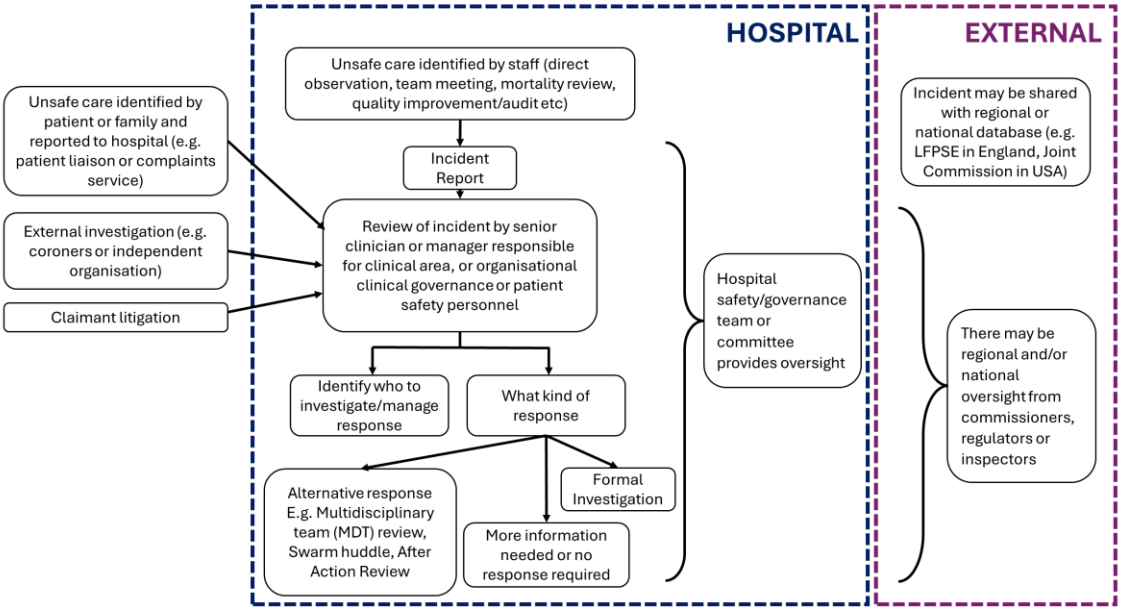
Table 1.2 Barriers and facilitators to incident reporting, ranked in order of frequency of reporting in the literature.[71]

Rank	Barriers	Facilitators
1	**Fear of adverse consequences** (such as litigation, blame, judgement from and towards colleagues, negative impacts of colleagues mental health)	**Organisational factors** (such as the provision of feedback/communication following incident reporting and a non-punitive incident reporting policy)
2	**Process and systems of reporting** (such as time required to complete an incident report and complexity of the reporting process)	**Process and systems of reporting** (such as reporting format, ensuring anonymity and/or confidentiality, and simplification of reporting)
3	**Incident characteristics** (such as level of harm, cause of incident and frequency of incident)	**Incident characteristics** (such as level of harm and frequency of an incident)
4	**Individual HCP characteristics** (such as a negative attitude/lack of value placed on incident reporting)	**Individual HCP characteristics** (such as a positive attitude towards incident reporting and a high value placed on incident reporting)
5	**Knowledge and skills** (such as lack of reporting clarity or clarity regarding what constitutes an adverse event)	**Knowledge and skills** (such as training in reporting)
6	**Work environment** (such as Workload/priority and accessibility)	**Team factors** (such as good teamwork/communication and a positive team culture)
7	**Organisational factors** (such as lack of feedback and communication following incident reporting and the absence/lack of a positive reporting culture)	**Professional ethics** (such as a strong sense of duty and responsibility)
8	**Team factors** (such as the negative impact that incident reporting could have on working relationships, the influence of seniors not to report, and how staff felt about reporting their peers)	**Work environment** (such as access to the incident reporting system and those whose workloads allowed for and those that prioritised incident reporting)
9	**Professional ethics** (such as a lack of personal responsibility to report)	**Fear of adverse consequences** (such as fear of litigation and fear of blame increasing the likelihood of reporting)

1.4 How are identified incidents responded to?

Once an incident is detected—whether through an incident reporting system, patient complaint or by other means—the details of the incident must be reviewed to determine an appropriate response. This process is briefly summarised in figure 1.2 below. The focus of this PhD is recommendations generated within hospitals, in response to incidents and investigations, but a brief outline of the reporting processes within hospitals and how they link with regional or national processes is outlined below. While the author is most familiar with the process within UK hospitals, the steps in the process are broadly similar in hospitals across the world,[63,65,69,70,91,92] suggesting some potential generalisability in the findings of this PhD.

Figure 1.2 Incident review process and selection of response following incident report



Within hospitals, the task of incident review is usually carried out by staff in clinical governance or patient safety roles, as well as those in clinical leadership or management roles (e.g. a nurse in charge of a ward).[70,93-98]

An appropriate response includes the following potential actions:

- Immediate action to reduce risk or harm to the patient involved in the incident, as well as other patients accessing the service in the immediate period following the incident.
- Disclosure of incident to the patient and or family/carers, which may include formal apology and information about the incident and planned response.
- Decision about what kind of further information gathering is required; whether in the form of a formal investigation or other approach, such as after-action review.
- Notification of other organisations, regulators, commissioners etc.

When deciding on an appropriate response, staff may consider harm severity and probability of re-occurrence.[94-96] Guidelines or policies suggest that those deciding on an appropriate response also consider the 'potential for learning' as well as degree of harm to the patient.[69,70]

Within hospitals, there is usually a committee that oversees the IRS and investigations. The committee is usually made up of a senior doctors, nurses, pharmacists and managers, who make decisions on incident management, approve investigations and are tasked with monitoring the effectiveness of the system.[69,91,92,99] There are also regional and national committees or bodies with oversight of these processes and reports, such as commissioners and regulators.[70,95-97,99-101]

Between 2004-2024, in England, incident reports from individual hospitals were anonymised and submitted to a national database, called the National Reporting and Learning System (NRLS), with the purpose of feeding back alerts and guidelines to organisations.[102] The NRLS was decommissioned in 2024 and replaced with The Learn from Patient Safety Events (LFPSE) service, with a similar role to the NRLS, in that it aims to analyse incident reports at a national level in order to guide improvement work at a national and local level.[102] Another system exists in England called the Strategic Executive Information System (StEIS), through which only serious incidents are reported for monitoring by commissioners and regulators within England.[69]

1.5 Introduction of the Patient Safety Incident Response Framework (PSIRF) in England

During the course of this PhD a significant policy change has occurred within the English NHS, in the approach to the management of incidents and investigations, through the introduction of the Patient Safety Incident Response Framework (PSIRF).[103] PSIRF represents a significant shift in policy, promoting a proportionate approach when responding to patient safety incidents by encouraging organisations (hospitals) to consider the balance of resource allocation between learning and the delivery of improvements.[103] The associated policy documents go on to state that:

"The PSIRF is not a different way of describing what came before – it fundamentally shifts how the NHS responds to patient safety incidents for learning and improvement. Unlike the SIF[Serious Incident Framework], the PSIRF is ***not*** *an investigation framework that prescribes what to investigate. Instead it:*

- *advocates a co-ordinated and data-driven approach to patient safety incident response that prioritises compassionate engagement with those affected by patient safety incidents*
- *embeds patient safety incident response within a wider system of improvement and prompts a significant cultural shift towards systematic patient safety management"*[103]

It remains to be seen, how these encouraged shifts in approach will be enacted by organisations and the subsequent effects on investigation practice and outcomes. The SIF was prescriptive about how serious incidents should be identified and managed, but did not provide guidance for those incidents not meeting the SI criteria.[69] In contrast, within PSIRF, the term serious incident has been removed and organisations are encouraged to take a systems approach, tailored to their priorities, to deciding which incidents are investigated and with which approaches.[103] For instance, an organisation may decide to carry out a formal investigation of a group of similar incidents, or carry out an after-action review following an incident similar to one that has recently undergone a formal investigation. The PSIRF policy documents describe this as a proportionate approach to incidents—allocating resource and deciding an learning

approach based on the organisation's context, priorities, previous incidents and expected outcomes or benefits of further investigation.[103] An After Action Review is described as a meeting in which stakeholders, or those involved in an incident, are brought together, for approximately 30-60 minutes, "to capture learning from to avoid failure and promote success for the future.".[104]

The introduction of PSIRF, may result in fewer formal investigations in the English NHS (Patient Safety Incident Investigations, PSIIs) in favour of alternative methods such as after-action-review or multidisciplinary team review. There will, however, continue to be PSIIs carried out when a more in-depth review of a single incident or cluster of incidents is deemed necessary. The PSIRF policy states that *"deaths thought more likely than not to have been due to problems in care"* will need to be investigated with a PSII.[105:p5] To date the types of incidents that have required a more formal investigation have included those that resulted in unexpected patient death or injury, actual or alleged abuse, and those incidents classified as 'Never Events' in England.[69,106] Never events are described by NHS England as "*safety incidents that are wholly preventable where guidance or safety recommendations that provide strong systemic protective barriers are available at a national level and have been implemented by healthcare providers*".[106:p6] A list of never events is available online, and examples include incidents relating to wrong site surgery or misplaced nasogastric tubes.[106:p4]

The policy also states there will be *"no further national rules or thresholds to determine what method of response should be used to support learning and improvement. Instead, organisations are now able to balance effort between learning through responding to incidents or exploring issues and improvement work."*.[105:p5] It could be argued that if fewer formal investigations are to be carried out, there is even greater need for these investigations, and subsequent recommendations, to be of high quality, and effective. These types of more formal investigations may be described as sentinel event investigations within other healthcare systems, internationally.[65,70] With this focus in mind, the process of incident investigations will be examined in more detail, but first the next section will consider some underpinning theoretical concepts.

1.6 Theoretical underpinnings of incident investigation systems

Since the landmark reports discussed in section 1.1, efforts to improve patient safety have relied heavily on the retrospective investigation of patient safety incidents;[3] an approach founded on an interpretation of safety theory which proposes that errors are multifactorial in nature and that identifying and addressing organisational latent failures (contributory factors) through investigation and generating recommendations, might reduce future recurrence.[87,107] Most incidents are due to human error, but reducing the risk of incidents requires a focus on the factors that make them more likely, the conditions in which people work, such as equipment, environment and supervision. This section will briefly consider the development of important theory and concepts, and models of accident causation, underpinning the approach to investigations within healthcare.

Before considering causation models, it would be useful to define the term 'system' in the context of healthcare. Pascale Carayon and colleagues suggest that *"healthcare systems can be conceptualized as work systems in which people perform multiple tasks using various tools and technologies in a physical environment and under specific organizational conditions".*[108:p3] Within healthcare, possibly more so than other safety-critical industries, there may be complex social interactions to consider, between patient's and families, patients and staff, and staff and staff. Furthermore, different organisations, teams, people (patients, family or staff) may have different and sometimes conflicting perspectives and goals.[109] The delivery of healthcare relies heavily on humans, and human-human interactions, more so than other more technical sectors. Shorrock described the idea that humans are often assisted by technology rather than technology assisted by humans.[110] It is also likely that within healthcare there are many contingencies, that are not fully predictable—goals, resources and contextual factors are dynamic and more changeable than those in other sectors.[109] These represent some important factors and challenges when consider what a 'system' comprises in the specific context of healthcare. In the next section the development of accident causation models will be discussed, which may be utilised when exploring incident causation.

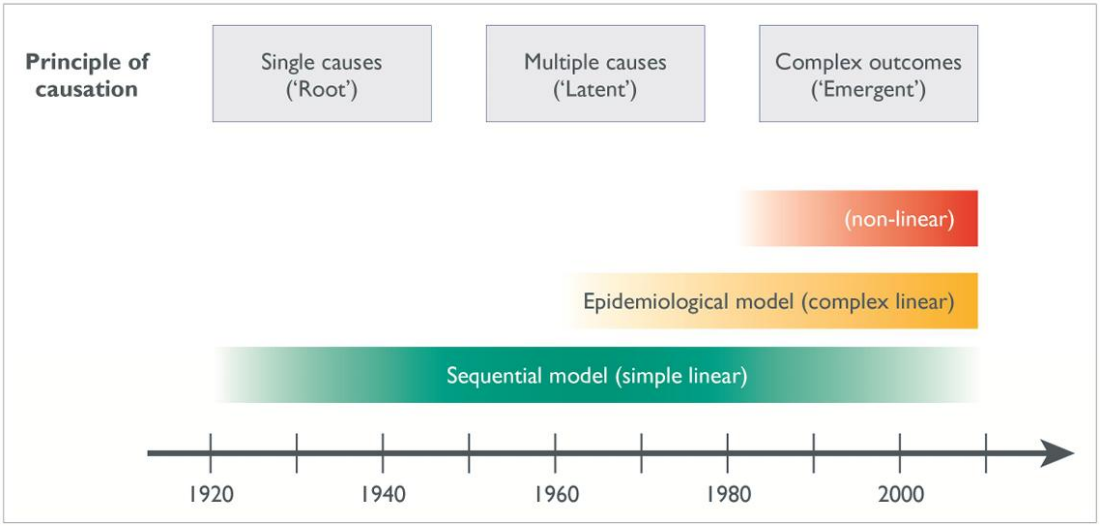
1.6.1 Accident Causation Models

Long before their adoption in healthcare patient safety, a succession of conceptual models of accident causation were developed, which can be divided into three distinct phases, as illustrated in figure 1.3. These phases have been called i) simple linear models (sequential); ii) complex linear models (epidemiological); and, iii) complex non-linear models (systemic).[111,112]

1.6.1.1 Simple & complex linear models

Simple linear models such as Heinrich's Domino Theory, suggests that accidents occur as a result of a sequential chain of events; invariably occurring in a fixed or logical order.[113] These early models proposed that accidents might be prevented by removing one of the factors, therefore interrupting the sequence leading to the accident.[113,114]

Figure 1.3 Summary of accident causation model development.[112]



Complex linear models, suggest that accidents occur as a result of a combination of unsafe acts, or human errors, by those in direct contact with patients or the system, and so-called latent failures within a system.[112,115] It would be useful at this point to consider human error in a little more detail, followed by a definition of latent failures.

1.6.1.2 Unsafe acts / human errors

James Reason defined human error as "*planned actions that fail to achieve their desired consequences without the intervention of some chance or unforeseeable agency*".[116:p18] These may take the form of failures of execution (slips, lapses, trips or fumbles) in which

the plan is good but the actions don't go as planned, or 'mistakes', in which the plan itself was not adequate to achieve an intended outcome, but the actions went to plan.[115] Mistakes can be divided into two types: rule-based and knowledge-based. Rule-based mistakes occur when 1) 'good rules', for instance evidenced-based guidelines, are misapplied to a situation, 2) 'bad rules' or incorrect procedures are applied, or 3) there is a failure to apply a 'good rule' such as follow appropriate guidelines. Knowledge-based mistakes occur when individuals have to improvise and problem solve in the moment, either because of a lack of knowledge or a lack of rules or routines.[115]

Violations are described as deviations from routines, rules or standards, which can be deliberate or unintentional, and further divided into three types: routine, optimizing, and necessary.[115,117,118] It is important to distinguish violations that are malicious and non-malicious. In other words, violations mostly occur without any intent to cause harm, or even represent an attempt to achieve a positive outcome.[115] For instance, healthcare staff might ignore a rule, originally designed to reduce risk, in order to expedite treatment for an illness, inadvertently placing a patient at increased risk.

1.6.1.3 Latent Failures

In these models, latent failures (also called latent conditions or factors) can be thought of as features of a system or organisation that might influence those working within it, providing the conditions in which active failures are more likely to occur.[116,119] For instance, a new hospital ward might be designed and built without an emergency alarm; three months after building is complete a patient on the ward becomes unwell, but without an emergency alarm staff are unable to quickly call for help. This might result in a delay to treatment for the patient and harm. Latent failures can be separated in time and space from the active failures that might follow, and this is why they are sometimes described as 'lying dormant'.[120] Other examples of latent failures include inadequate systems of communication, inappropriate skill mix of staff on a ward, the design of policies or procedures, arrangements for supervision of junior staff.[119] A number of frameworks, attempting to identify latent failures have been developed,[121-125] and the empirically developed Yorkshire Contributory Factors Framework is discussed below in section 1.6.2.

A detailed description of the development of complex linear models and important work and researchers in this area, for accident causation, is beyond the scope of this thesis. It is however important to highlight how accident causation models have developed as they underpin the process of investigations which inform the generation of recommendations. Through successive iteration, complex linear models have identified some important considerations, such as the need to move away from identifying simple causal explanations, to exploring how multiple factors within a system combine to result in accidents.[115,126] Complex interactions, separated in time and space, between elements of socio-technical systems, combine to result in accidents. The analysis of these accidents may only present a record of one past event which may not capture the dynamic and unstable system in which the incident occurred.[116,127,128]

Much of accident causation modelling work was distilled in and informed the development of the organisational accident model (OAM), which will be presented below with an example to illustrate how it informs the investigation of incidents in healthcare.

1.6.1.4 Complex non-linear models

Complex non-linear models represent an evolution of accident causation modelling that highlights complexity. Leveson described systems as *"interrelated components that are kept in a state of dynamic equilibrium by feedback loops of information and control"*.[129:p250] Perrow argued that the components in a system (such as people, environments, technology) are linked by multiple channels, might affect each other unexpectedly or unpredictably and fully understanding these interactions might be impossible.[130,131] Hollnagel has contributed greatly to the field of accident modelling and safety;[112,132-135] The Functional Resonance Accident Model (FRAM) is an important move away from linear sequential models. FRAM highlights the variability within systems, not as a good or bad thing, but simply a result of complex systems. It promotes a shift away from simply identifying the 'causes for an incident', to understanding how systems function and vary, how they achieve their intended goals, how conditions leading to an incident may emerge.[134]

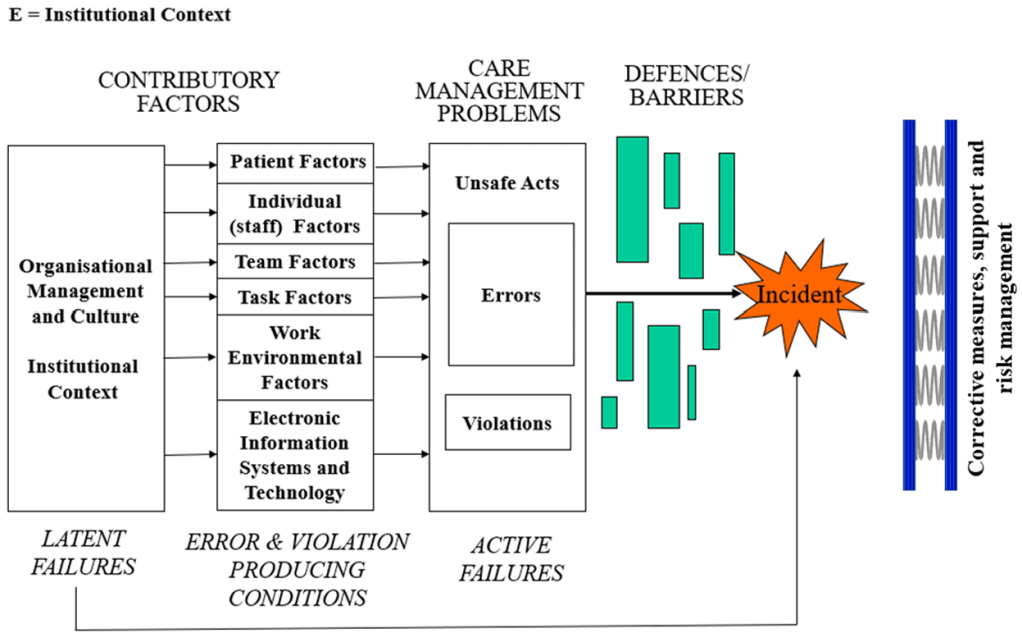
While the accident analysis domain has advanced significantly,[136] there are some important healthcare specific critiques to highlight at this point. Firstly, as discussed in more detail below, healthcare investigations have predominantly drawn on linear models such as the organisational accident model.[98,115] Only recently have more developed models such as the Systems Engineering Initiative for Patient Safety (SEIPS) been encouraged within healthcare policy. While this adoption is encouraging there is no guarantee that more advanced models will be effective in supporting system-based learning, unless further research is undertaken to adapt these to the healthcare context.[136] The second critique relates to the emergence of the safety-II approach.[137] Safety-I is concerned with avoiding failure to reduce harm, in contrast to safety II which encourages the "study of how people and systems are able to succeed under variations so that the number of intended and acceptable outcomes is as high as possible".[137,138:p97] The investigation of incidents is considered to sit within the safety-I approach and therefore argued to miss the opportunities for improving safety offered by the safety II approach. While debate continues about how safety-I and -II should be effectively applied in the context of patient safety, it is likely that there is merit in utilising both approaches.[138,139] Suffice to say, investigations in healthcare have historically adopted a safety-I approach and this will be explored in greater detail in the following sections.

1.6.1.5 The Organisational Accident Model

The key purpose of outlining the development of accident causation models, although briefly, is to demonstrate the theories that underpin the activity of incident investigation which informs the development of recommendations. Incident investigation in healthcare, to date, has largely drawn on the organisational accident model (OAM), as illustrated in figure 1.4.[98,115] Despite the development of accident causation and patient safety theory and models, as illustrated above, beyond the OAM, healthcare's adoption of new approaches to investigation has largely stalled, and failed to move beyond the OAM and RCA.[12]

For those less familiar with incident investigations in healthcare a fictional example incident and demonstration of how the OAM might be used to help understand how an incident occurred is included in appendix 1(page 194).

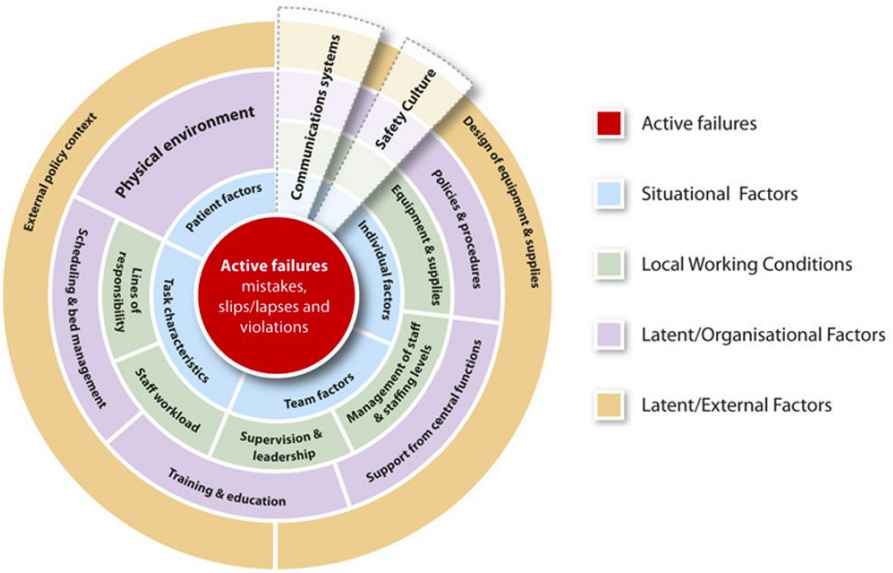
Figure 1.4 Extension of James Reason's Organisational Accident Causation Model from the London Protocol.[98,115,140]



1.6.2 Identifying contributory factors

An important part of an investigation, guided by the OAM, is the identification of factors, within a system, that have or may contribute to unsafe care. A number of structured frameworks have been adopted within healthcare to support investigators in identifying contributory factors.[121-125,141] Many of these have been originally developed within non-healthcare sectors, without significant adaptation or empirical research, potentially reducing their applicability.[142,143] Lawton and colleagues drew on the growing literature base on contributory factors in healthcare incidents to develop the Yorkshire Contributory Factors Framework, illustrated in figure 1.5.[144] This hierarchically ordered framework presents categories of CFs, as described by the authors, from "proximal (sharp end) to distal (latent)".[144] In this way the framework conceptually aligns with the OAM described above, in identifying that the active failures of individuals are usually influenced firstly by proximal, in time and space, 'situational factors', such as fatigue, distraction, complexity of a patient case, followed by increasingly more 'distal' factors from decisions about staffing on a ward, to the design of a ward layout, design of equipment and even national health policy.

Figure 1.5 The Yorkshire contributory factors framework.[144]



Lawton and colleagues highlighted that across the 95 studies reviewed, the majority of CFs identified by investigations were active failures or individual factors.[144] This tendency was highlighted again in a subsequent study by Peerally and colleagues, using the Human Factors Analysis and Classification System(HFACS), an alternative CF framework.[145] It has been suggested that the design of these influential frameworks themselves may encourage a focus on active failures; Lawton highlighted that AIMS contains 33 codes for human factors and 21 to system,[121] while the Eindhoven classification has nine for human failure, four for technical and five to organisational failure.[122] Peerally and colleagues identified, using the HFACS framework, that 6% of investigations reported CFs that *"lay beyond the remit of the Trust",* described as extra-organisational factors.[145:p.430] One such factor was the national shortage of staff with particular skills. This study highlights not only the importance of these frameworks in guiding the identification of CFs but the importance of continued work to empirically develop and iterate the frameworks.

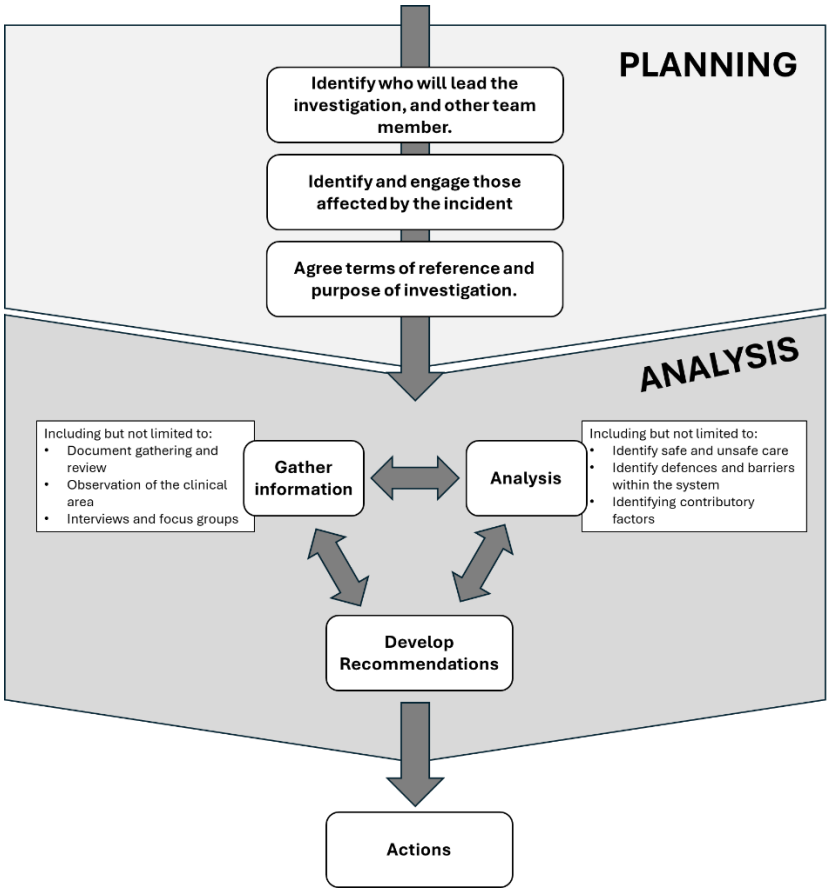
The preceding sections have explored the scale and impact of harm and the theories underlying how patient safety incidents might occur. The next sections will consider the systems that have been set up in healthcare to improve patient safety, which have largely focused on incident reporting and investigation. The focus of this PhD is the recommendations following investigations. However, to explore recommendations it is important to consider the preceding steps, namely: 1) detection/reporting of incidents,

2) selection and coordination of response or investigation method, 3) approach to incident investigations. Each of these will be explored in more detail in the following sections.

1.7 Patient safety incident investigations in practice

While a range of tools are available to support investigators in understanding the factors that contribute to incidents (briefly discussed below), investigations generally follow the steps illustrated in figure 1.6. However, there are variations in the exact terms and steps described within available guidelines and publications,[69,70,98107] and so I present a version that integrates these and my own experiences (as a doctor in acute medicine and patient safety investigator) and interpretation(Figure 1.6), as well as define important terms, to frame the remainder of this thesis.

Figure 1.6 Overview of the 'usual' investigation process in hospitals[69,70,98,107]



Following the decision to investigate, there are a number of considerations:

1) Identification of investigation lead, and the formation of a team of at least two people to conduct it
2) Identification of those affected by an incident (such as patient, families, carers, staff), who need to be supported and engaged and involved in the investigation process
3) Consideration of the scope and purpose of the investigation
4) Gathering information such as patient records, documents, interviews, observations of clinical practice
5) Exploration of the incident that triggered the investigation, which might include establishing a chronology of events
6) Analysis of data and information to identify unsafe care, risks, factors that contributed to the incident, as well as usual practice and factors that might mitigate or reduce risk
7) Generation of recommendations to guide where improvements might be made to ensure patient safety and quality of care (the terms recommendation and action are defined below)

It might well be that an investigation is designed to consider a cluster of similar incidents, sometimes called an aggregated review. While the suggested steps of a formal investigation (PSII) have not significantly changed from SIF to PSIRF, aggregated review is strongly encouraged.[103]

The above steps do not necessarily occur sequentially or only once, but rather the investigation is likely to proceed via a process of iterative exploration and theory development. For instance, analysis of information may indicate that more information or data is needed, or that other stakeholders need to be involved. Vincent described incidents and the subsequent investigations as a '*window on the system*', an opportunity to *"reflect on what the incident reveals about the gaps and inadequacies in the healthcare system in which it occurred"*.[146:p242]

Root Cause Analysis (RCA) has been widely promoted as an approach to conducting healthcare investigations since the 1990s, originally developed within high risk

industries such as nuclear, aviation and aerospace.[12,69,147-150] In the USA, RCA was introduced into healthcare in 1996, based on its use in the NASA shuttle program,[151] and has continued as the predominant approach referenced in policy and academic literature. The exact process and application of RCA varies between sectors and organisations, but the basic premise is to encourage an approach to identifying systemic causal factors for deficiencies in care. RCA is not based on a specific theoretical framework, but rather comprises a 'toolbox' of techniques such as management oversight risk tree, barrier analysis and cause and effect charting purported to aid in the identification of causes for an incident or problem.[107] It is somewhat surprising that despite the large body of accident causation research, summarised in 1.6 section, and approaches to incident investigation based on safety theory available (such as STAMP, AcciMap, HFACS and the London Protocol), that RCA is so ubiquitously used in healthcare.[152-155] RCA could be argued to be too linear and simplistic a method for exploring the complex systems in which healthcare is delivered, and its effectiveness in healthcare has come under significant criticism.[12,156-158] RCA and methods for investigation will be discussed later in the thesis.

A key element of an investigation is the generation of recommendations, which is the focus of this PhD. In the final part of this chapter a working definition of recommendations will be presented as well as an important distinction between recommendations and actions. A more detailed discussion of the generation of recommendations and critique will be presented in chapter 2.

1.8 A working definition of a recommendation

Clarity on two key terms (recommendations and actions) will be discussed here, providing a working definition for the remainder of the thesis. A key purpose of investigation has been to prevent incident recurrence, through the generation of recommendations for changes or improvements.[124,156,159] There is often confusion between recommendations and actions,[160] so to ensure clarity throughout this thesis definitions for both are provided:

Recommendations set out what improvement is needed, without defining how that improvement is to be achieved. Recommendations allow opportunity for a range of

actions. Recommendations are linked to system performance such that the reason for the change remains understood as the solution is developed and implemented.[160]

In the English NHS, PSIRF uses the term 'areas for improvement' instead of 'recommendations' to *"reduce the likelihood of solutionising at an early stage of the safety action development process",* "Areas for improvement set out where improvement is needed without defining how that improvement is to be achieved". [161:p.2]

Actions are concerned with satisfying recommendations in a way that is practical, effective and sustainable. Actions are developed and tested. Not all actions will be implemented. The solution eventually agreed and implemented will depend on resources, cost and measurement of effectiveness.

Chapter 1 Summary

The field of patient safety is relatively young in comparison to the practice of medicine, and only within the last few decades has the full extent of patient harm been realised. Efforts to improve patient safety and reduce harm have centred around the reporting and investigation of incidents, followed by the generation of recommendations. While investigations themselves have drawn on broader safety theory, this has focused on the organisational accident model rather than continuing to follow subsequent developments in safety science. The next chapter will consider the recommendations generated from investigations, alongside reasons why this became the focus of this PhD and methodological considerations.

Chapter 2 will discuss:

- why this PhD was undertaken
- experiences and observations of safety in a hospital
- positionality and epistemological perspective
- initial exploration of the problems with recommendations
- **Aims and objectives and rationale for the methods chosen to answer these.**

Chapter 2 Developing and designing this PhD

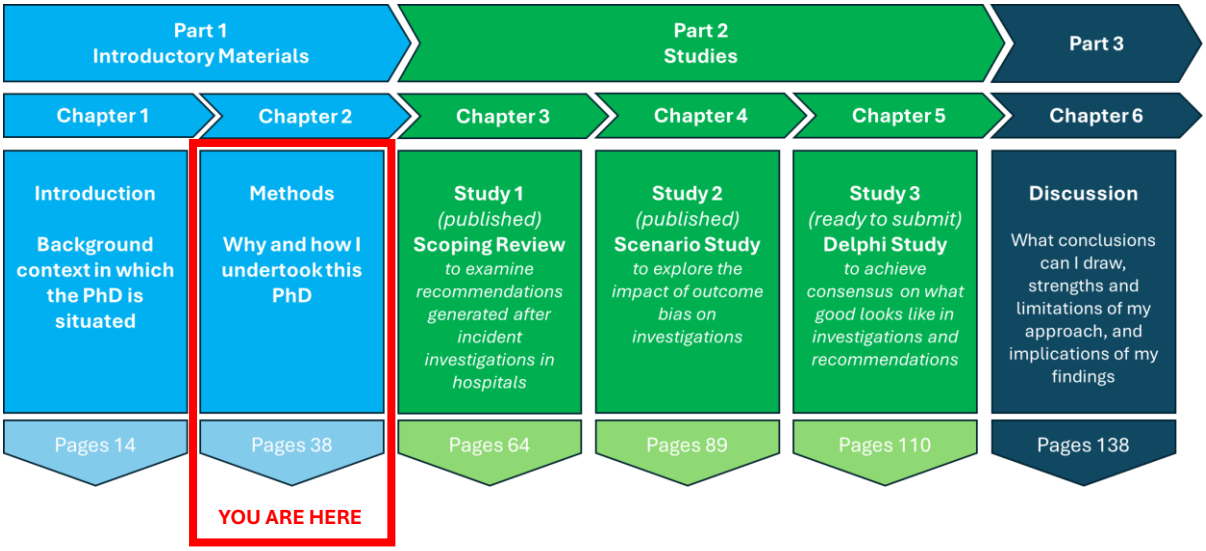
Chapter 1 presented some important context and background in which this thesis was situated. Chapter 2 presents the path of inquiry that led to my PhD and this thesis.

Section 2.1 presents reflect on the authors journey into patient safety research, presenting experiences from an early career mistake, front line patient safety, and undertaking a hospital incident investigation.

Section 2.2 presents a discussion of the authors positionality and epistemological perspectives and how these shaped the approach of the PhD.

Section 2.3 summarises the initial exploration of the problem of the PhD and thesis, namely the recommendations following incident investigations within hospitals. This section is split into observations from a local hospital perspective, followed by a review of the literature on recommendation generation.

Finally, the aims and objectives will be presented in section 2.4 and the methods used to address these in section 2.5.



2.1 My journey into patient safety research

My motivation to undertake this PhD largely came from being a hospital doctor, who was becoming increasingly aware of the complexity of how care was delivered and wanting to explore the possibilities to improve safety. On the one hand my motivation was very personal – as a doctor I wanted to be part of caring for patients in a safe and effective manner, desperately wanting to avoid causing harm. I wanted to be *a 'safe' practitioner*. The other motivation was less personal, wanting to explore how the healthcare system – made up of hospitals, GP practices, pharmacies etc – delivered safe care. I wanted the *system to be safe* for both patients, and for me to practice within.

Looking back there were many experiences that likely led me to undertaking this PhD in the way I have. However, for brevity I have chosen three examples, described below.

2.1.1 My own practice and my own 'mistakes'

Around 10 years ago I made a mistake that has stuck with me ever since. I had seen a patient during the day and had requested that a chest x-ray be done to look for any evidence of air under the diaphragm. This would have suggested that the patient had a perforated bowel. This is a serious problem that might require urgent surgery as it could lead to serious infection and death. The chest x-ray had not been done before the end of my shift and so I should have handed this task over to the doctor coving the night shift. I forgot to do this. I came back to work in the morning to find that the chest x-ray had been done and did show air under the diaphragm but that no one had looked at it or realised until the morning. I remember the feeling of dread and worry. Thankfully, the patient was well; and after realising the delay in reviewing the x-ray, they were promptly reviewed by the surgical team who advised no surgery. He was discharged home a day or two later.

I remember speaking to one of the senior doctors, asking *"what I should do? Should we report the incident? How was I going to make sure this didn't happen again?"* They didn't seem too concerned *"Don't worry. These things happen, the patient's fine."* That was the end of the matter.

There are probably many reasons this experience has stuck with me. I had lots of personal concerns such as worry about what would have happened if the outcome for patient was different; what if they had died or come to harm? Could I be sure I wouldn't forget anything again? I also started to think more about the wider system I was working in – what was it doing to 'help' me, how did it support humans, with their fallibilities, to be safe? Additionally, this was an incident, but no investigation occurred, no attempts to avoid a repeat. So important an experience this was for my own personal journey into research, that an anonymised version of this incident story was used within the scenario study of this PhD.

I have made other mistakes since this incident, fortunately not resulting in harm to patients (as far as I'm aware), and as a human, will continue to make mistakes. It is this knowledge that continues to drive my enthusiasm for patient safety research.

2.1.2 Front line patient safety

Following the experience above, my interest in quality improvement and patient safety grew. I got involved in projects, sought out training and met with safety and quality teams in the hospital I was working in. I learnt about incident reporting and investigations. I attended regular serious incident group meetings, during which investigation reports were reviewed and discussed. I was encouraged to see the effort and time staff put into improving safety. A number of challenges were clear in terms of staffing and resource. A relatively small number of people, two to three, were responsible for overseeing an incident reporting system that might receive over 15,000 reports in a year. Staff had to find time alongside clinical or operational duties to review reports and undertake investigations. Some investigations might take weeks to months to complete. There was then a great deal of work and resource required to action recommendations following investigations.

I wanted to know more about patient safety and investigations and, as a doctor, used to evidence-based medicine, I wanted to know about the research that guided patient safety.

2.1.3 Leading an investigation

My final reflection concerns an investigation that I was asked to lead. A patient had died in hospital and an incident report had been submitted. The incident report and initial review identified that there may have been a number of issues with care, such as delays in review, nutrition and response to clinical deterioration. I learnt a lot doing this investigation, including, but not limited to:

1) The difficulty of balancing the time required for a thorough investigation alongside other duties, and sometimes guilt that I was not able to commit more time to the investigation.
2) That finding time to engage with staff, either directly involved in the incident or working in relevant areas, was difficult due to clinical work pressures for them.
3) For staff, patient safety and investigations could be a highly emotional topic. Staff were fearful of blame and repercussions, and it seemed that this was a barrier to honest discussions.
4) That it was hard to know how to engage family members in the investigation, and a distinct lack of available guidance on this – from a local or national policy perspective.

While I learned a great deal by undertaking this investigation, it left me with many more questions about how and when investigations should be done. It felt like the most important element was the recommendations that I would make at the end of the investigation. Based on what I had found, what would I recommend that the hospital do to improve safety or prevent recurrence of this incident? I realised by doing this investigation and trying to develop recommendations that not only did I not have the knowledge and skills needed, but that there was little guidance available for any investigator. I needed to find out more, I needed to get involved in research in this area.

During the course of the investigation, I came to the conclusion that although I did identify episodes of unsafe care, I was not convinced that any of these directly resulted in significant harm to the patient, or caused their death. I did however make a number of recommendations relating to system issues that I considered important, things that if addressed may prevent future unsafe care or harm. When I presented the findings of my

investigation, due to my conclusion that the unsafe care did not directly result in significant harm, the SI was downgraded. This meant that it was no longer considered an SI. This immediately removed the need to urgently action the recommendations. I remember being disappointed by this and also concerned that the issues I had identified may go unaddressed. It seemed strange that at this point, after my investigation, the harm was given such importance; I had identified unsafe care and system deficiencies that needed addressing.

While I've described a select few above, many experiences have contributed to me undertaking this PhD, maintained my focus and shaped its course. In the following section I will move from motivations to consider my positionality and epistemological perspective in relation the research.

2.2 Positionality and epistemology

I have described above some examples of my motivations to undertake this PhD. It is increasingly common, particularly within a thesis, for researchers to consider concepts of identity, positionality and epistemology and how they might impact the design, analysis and interpretation of research.[163,377-380] In this next section, I first describe how my professional identity has changed during the course of this PhD, followed by a reflection on my positioning between academia, clinical practice and frontline patient safety. Finally, I will describe my epistemological perspective in relation to the research I have undertaken within this PhD.

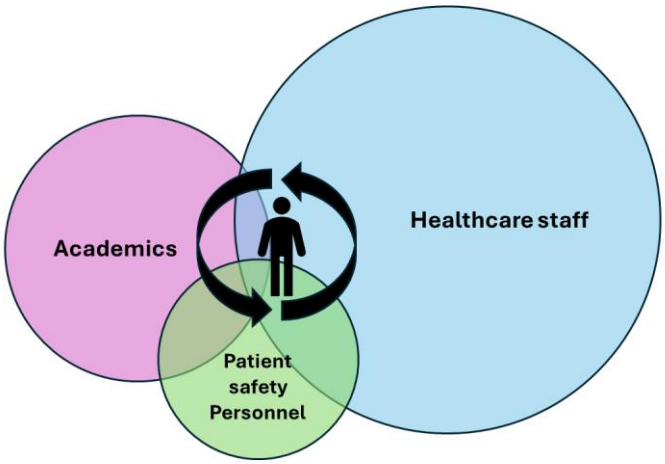
2.2.2 Positionality

Positionality pertains to my place within this research, and the interactions between my personal background, my values and perspectives, and how these influence the studies and course of this PhD.[377,381,382] Holmes describes three areas to consider in terms of a researchers' position in relation to: 1) the topic under investigation, 2) the research participants, and 3) the research design.[383] Further, positionality is not only impacted by where I place myself within the research, but also where others place me, such as supervisors, collaborators, and research participants.[162]

There are a number of concepts and terms used to describe an academic's position in relation to the community or context in which the research is being undertaken: **insiders** undertake research from within the community, while **outsiders** study from outside, and **in-betweeners** are considered neither completely inside or outside of the studied group.[163-165] Others argue that positionality is less rigid and represents a continuum along which academics move back and forth.[162]

On reflection of my own positionality, represented in figure 2.1 below, I would relate most to the idea of a continuum. However, I recognise that during my PhD there have also been multiple 'groups' that I have moved between.

Figure 2.1 My research positionality



The first group is *healthcare staff*. I am a clinical doctor and I work with healthcare staff in the hospital environment every week. My PhD has been part-time and so I have maintained my clinical practice alongside my research. During the Covid-19 pandemic I paused my research to return to full time clinical work; this period allowed me unusual insight into healthcare systems and how they reacted to such change.

The second group is *patient safety personnel*; these are staff within the hospital who have formal patient safety roles such as incident investigators or patient safety or risk managers. I have worked with staff in these roles and joined committees or working groups looking at topics such as incident investigations, incident reporting and medication safety. Over time I have engaged with a wider network of patient safety personnel from hospitals across the UK and abroad. My time and connections within

the healthcare staff and safety personnel groups provides me insights into the system, how it works and direct observation and experiences. Professional relationships within these groups allow me a level of trust and the opportunity for open and honest discussions about safety. It also allows me the opportunity to discuss theories, ideas with key stakeholders as well as disseminate my findings and research. These connections have been highly beneficial in forming collaborations for research as well as recruitment of participants.

The third group is *academics*. Certainly, at the outset of my PhD this is the group that I least related to, and felt like an outsider within. I have found that over time and through the course of my PhD I have felt increasingly comfortable within this group. I have found my clinical and frontline safety experiences have been of particular interest to researchers who might be described as outsiders,[163] and have provided some common ground to build professional relationships.

2.2.3 Epistemological perspectives

The term epistemology is described as the study of knowledge,[166] and defined by the Oxford English Dictionary as "the theory of knowledge and understanding, especially with regard to its methods, validity, and scope, and the distinction between justified belief and opinion." It is described as being concerned with how we, as humans, acquire knowledge and understanding about the world around us.[167,168] It is therefore suggested that researchers should consider and reflect on their epistemological perspective or approach,[169] as it may explain what "kind of knowledge is considered truthful or valid, false or coincidental.[170]

This has particular relevance to this PhD in that I aim to examine or investigate the activity of investigation of incidents in healthcare. Investigations in healthcare, themselves, aim to understand how incidents occur, to formulate some understanding of how components in a system interact to produce outcomes for patients. While I did not intend to explore the epistemological approach of incident investigators, it is at least important to acknowledge the need to consider my approach to knowledge formulation. There are a number of epistemological perspectives or positions, or ways

in which people might believe truth exists or knowledge is acquired, such as idealism or relativism.[170]

Each epistemological approach has its limits in application, and it is likely that researchers will adopt different approaches when considering different questions and problems.[170] Pragmatism is an epistemological perspective that can be described as cutting across often dichotomous positions, encouraging a flexible approach and mixed methods to focus efforts on solving practical problems within the real-world.[171,172] Kelly and Cordeiro suggest the pragmatic approach is particularly useful when looking at organisational processes, allowing for exploration of connections between knowledge and action.[173] They go onto discuss three methodological principles underpinning pragmatic inquiry, namely 1) emphasis on actionable knowledge; 2) recognition of the interconnectedness of experience, knowing and acting; and, 3) inquiry as an experiential process.[173] The activities of incident investigation and recommendation generation are themselves highly complex socio-political processes,[159] with investigators tasked with making sense of highly complex socio-technical systems in incidents occur.[147,156,174,175] As a clinician and someone involved in front-line patient safety I was driven to produce actionable findings from research. For these reasons I have taken a pragmatic approach to this PhD. The process of incident investigation and generation of recommendations is a largely human driven process and therefore the pragmatic approach, using human experience as a way of building knowledge and understanding, is an appropriate lens.[176,177]

This thesis is anchored in the experiences of patient safety personnel and investigators, including myself, as a way of ensuring practical relevance. The continued involvement and feedback from those involved in investigations and the generation of recommendations, maintains the orientation of the research to real-world problems and actionable research. Throughout my studies I have attempted to triangulate and explore the complex nature of recommendation generation. Alongside research studies I have immersed myself within the organizational investigation processes with hospitals.

2.3 Initial exploration of the problem

Chapter 1 has provided some of the background context to this PhD, the first section of this chapter has presented the authors motivations, identity, and philosophical approach to research. The following sections will look at the problem that this PhD examined, namely the generation of recommendations following incident investigations. This PhD was shaped by the authors personal experiences of safety and the literature, both of which will be presented in the following sections.

As a result of the authors positionality as a clinician first, and academic second, the problem was initially explored from the perspective of the service – the current practice within a hospital. The following sections will illustrate the path of inquiry that led to this PhD. This began with examining recommendations following investigations within one hospital alongside broader observations of frontline safety practice. This initial inquiry led to a number of questions, to which the author turned to the literature for answers. It is at this point that gaps in the literature were identified, which formed the basis of this PhD, and shaped the thesis aims and objectives.

2.3.1 Exploration of the local recommendation generation process with an acute hospital trust

As a resident doctor I became involved in patient safety and quality improvement activities within the organisation that I was working clinically, in around 2014-2015. I worked with the hospital's patient safety team and attended meetings in which serious incident investigations were reviewed (Serious Incident Group). I was a member of a medication safety committee that reviewed medication related incidents and investigations and chaired a multidisciplinary group to evaluate the organisations incident reporting system.

In this section I will present some of my observations, but first for context, provide some information about the organisation I was working in. The organisation was made up of three acute hospital sites and five community hospitals, providing healthcare for approximately 800,000 people living in North Yorkshire. The organisation employed over 10,000 staff, and I was mainly based at York Hospital which had over 700 inpatient beds and provided a range of inpatient and outpatient services.

2.3.1.1 Organisational process

Patient safety incidents were reported via the online risk management system within the Trust, Datix. A team (patient safety & governance team) reviewed incident reports on a daily basis, and if there was a concern that an incident might meet the definition of a serious incident (SI), then details were presented at a weekly meeting of the Quality & Safety Group (a multidisciplinary group of senior clinicians and managers), to decide whether a SI investigation should be undertaken. The definition of a serious incident used was as follows:

"Events in health care where the potential for learning is so great, or the consequences to patients, families and carers, staff or organisations are so significant, that they warrant using additional resources to mount a comprehensive response. Serious incidents can extend beyond incidents which affect patients directly and include incidents which may indirectly impact patient safety or an organisation's ability to deliver ongoing healthcare."

If an incident was declared an SI, the Patient Safety and Governance team reported this via the Strategic Executive Information System (STEIS), NHS England's web-based serious incident management system discussed in chapter 1. An investigator was identified and it was expected that the investigation would be completed within 60 days. The investigation was usually undertaken by a single senior clinician, who would summarise their findings and recommendations within a report. Investigators were provided with a report template with the following sections:

1. Executive summary
2. Incident description and consequences
3. Pre-investigation risk assessment and initial determination of root cause
4. Terms of reference
5. The investigation team
6. Investigation type, process and methods used
7. Being Open with Patients
8. Involvement and support provided for staff involved
9. Detection of incident

10. Notable practice
11. Contributory factors to be addressed by Directorate
12. Any Immediate Actions taken as a result of the incident
13. Root cause
14. Likelihood of repetition
15. Arrangements for shared learning
16. Appendices
17. Recommendations and Action Plan
18. Action Plan Risk Assessment

Once the investigation was complete, the report would be discussed at a meeting of the SI group (multidisciplinary group of senior clinicians and managers). During this meeting, to which the investigator would be invited, the group would have the opportunity to review the findings of the investigation, the contents of the report and recommendations. The SI group might suggest changes to the report and/or recommendations.

Specifically for recommendations, the report would need to include the following details:

1. Recommendations (linked to findings)
2. Actions to achieve recommendations
3. Lead Responsibility
4. Date action agreed with Lead
5. Timescale and audit plan
6. Recommendation Evidence

2.3.1.2 Processes for recommendation generation

There were no details within the reports, to describe how recommendations were generated. While recommendations were critiqued during SI group meetings, there was no discussion of specific methods for generating recommendations or specific criteria for judging their effectiveness. It was often unclear how recommendations specifically related to the findings or contributory factors (CFs) identified in the report, or how recommendations had been arrived at.

As part of my work with the patient safety team in the Trust I was asked to review the CFs identified within investigations. I was provided with a set of 40 SI reports and asked to identify CFs included within the reports, as classified by the Yorkshire Contributory Factors Framework (YCCF) presented in chapter 1 (section 1.6.2). A total of 115 contributory factors were identified, which equated to approximately three per investigation. 82 of the contributory factors fell into the proximal categories within the YCCF. In other words, 71% of identified contributory factors focused on the active failures of individuals, or situational factors and local working conditions. 33 of the contributory factors could be categorised as latent or organisational, and none fell into the 'external factors' domain. While it was unclear, from reports, how investigations actually moved from CFs to recommendations, it seemed important that the investigation findings focused on individuals and active failures, with less exploration of wider system factors. This observation became more important when identifying a similar pattern in the types of recommendations generated, as discussed in the next section.

2.3.1.3 Recommendations proposed

The reports contained sections for both recommendations and actions but, referring to the definitions of both of these terms in section 1.8, there was overlap and confusion in the reports. In some reports recommendations appeared to be actions, in others actions read as recommendations, and in some reports statements labelled as recommendations or actions were actually neither. Investigations typically presented two to six recommendations each. A significant proportion of recommendations focused on reminders, the generation of a new or updated policy, or training. When considering the recommendation against the action hierarchy, presented below in section 2.3.2.2, most of them would be categorised as weak, with a minority considered medium or strong.

Following this review of SI reports, and based on my experiences of patient safety described earlier in this chapter, I began to formulate a number of key questions about investigations and recommendation generation:

1) How are recommendations generated? What tools are used?

2) What makes a 'good' recommendation and how is this defined?
3) Do recommendations for improvements focus on individuals, and why?

It was at this point that I turned to the literature, in an attempt to answer the questions above.

2.3.2 What is the evidence for the process of recommendation generation?

The generation of recommendations after incident investigations, has in recent years come under increasing academic scrutiny.[10-12,15] The following section will present several studies or publications which provided some insight into the key questions I had formulated.

2.3.2.1 How are recommendations generated? What tools are used?

Reviews by Dückers in 2009 and Card 2012 (examining 38 and 60 studies respectively), found a surprising lack of description of how recommendations were generated within hospitals.[10,178] While studies have been published suggesting ways in which recommendations could be generated,[179-181] there seems to be a gap in the literature in respect to how they are generated, in practice, within hospitals. In addition to the lack of empirical work examining recommendation generation, there is a lack of practical guidance on how to generate recommendations.[15] However, it is important to note that these issues extend beyond health care. Evidence suggests that a lack of guidance – as well as a plethora of other sociotechnical factors – impedes the generation, implementation, and evaluation of recommendations across safety investigations in other safety critical contexts such as rail, maritime, and nuclear.[9,15]

2.3.2.2 What makes a 'good' recommendation and how is this defined?

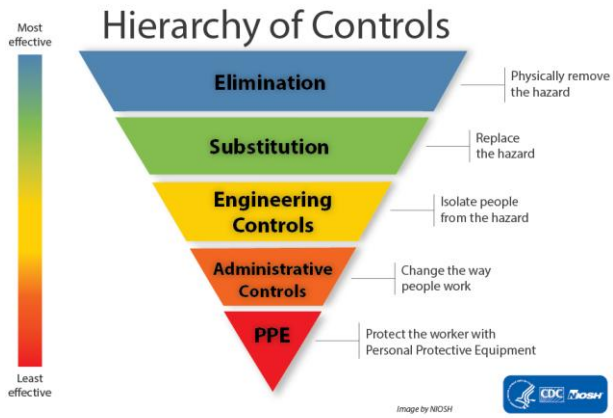
A number of studies or reviews have adopted the action hierarchy, discussed below, in order to evaluate the perceived effectiveness of recommendations.[10,100,182] The judgement of recommendations by both Card and Hibbert represent secondary analysis, i.e. not performed by the hospital or those that carried out the original incident investigations, but rather as part of a retrospective research study. Card reported, in his review, that the hospitals conducting the investigations provided almost no information on 1) how they judged recommendation effectiveness; 2) data on implementation rates;

and, 3) measurements of success or follow up.[10] Thus, while there is evidence in the literature of how recommendations might, in theory, be judged for effectiveness (often on presumptive effectiveness according to the action hierarchy), there is little about how hospitals actually do this. Further, few studies have examined the issue of recommendation 'quality' or effectiveness in the healthcare context.[12] Therefore, there is a gap in the academic literature and practical guidance as to how recommendations should be judged for effectiveness or quality.

The 'Action Hierarchy' (AH) or 'Hierarchy of Controls', as its use is widely referred to within the literature, is often used to make a presumptive judgement about recommendation quality or strength.[10,97,100,101,182,183] For this reason it is important to the context of this thesis and recommendation generation more broadly. The previously mentioned confusion between recommendations and action is evident in the use of the AH in healthcare; despite being called the 'action' hierarchy it is often used to examine both actions and recommendations.

The AH, developed outside of the healthcare context, is a hierarchical order of action types designed to control or eliminate risks, illustrated in figure 2.2.[184,185]

Figure 2.2 The Hierarchy of Controls (https://www.cdc.gov/niosh/hierarchy-of-controls/about/index.html)



The action types are rated on their effectiveness or 'strength' based on their presumed ability to reduce or eliminate risks. The strongest actions eliminate risk, for instance removing a dangerous element from a work environment. The hierarchy would suggest that the weakest actions depend more on human intervention, such as reminders or

training. The AH attributes a greater impact to action plans that significantly modify processes or systems.[186]

While the adoption of the AH in healthcare is widespread there is variability in its application, format and contents.[10,100,101,182-184] The AH will be considered through the remainder of this thesis.

2.3.2.3 Do recommendations for improvements focus on individuals, and why?

The review by Card and observational study by Hibbert both identified that recommendations focused on reminders, policies and training, with a disproportionate focus on individuals.[10,100] A number of reasons have been suggested for the poor quality of recommendations, or their tendency to focus not on addressing latent failures of the system (eg, design of equipment), but the behaviours of individuals, such as reminders, writing or rewriting policies and (re-)training staff. While studies in this area are limited, they suggest a lack of investigator expertise and limited guidance as important issues,[10,156,179,180] as well as poor feedback loops,[187,188] and the potential impact of a blame culture.[159,189]

Another potential explanation mighty lie within the field of psychology, and specifically relate to investigator bias. There has long been evidence that our judgements of individual responsibility or culpability are driven by the outcome of an accident or adverse event.[190,191] The same behaviour or actions are judged more harshly when the outcome is bad, than when the outcome is not.[190] In Walster's important study participants judged the same behaviour within a fictitious scenario (parking a car on a slope without putting the handbrake on) more harshly when the outcome was bad (the car moves and runs someone over), than when the outcome not (the car does not move or moves but no-one is harmed).[190]

Studies in healthcare have demonstrated that judgements of staff actions and behaviours, are also influenced by what is known as 'outcome bias'.[192-195] Outcome bias involves evaluating an individual or procedure responsible for an outcome; the evaluation is considered biased when outcome information is given excessive weight.[196,197] The issue of bias has largely been ignored within the policy and practice of healthcare incident investigation—that those investigating, or even consulted as part of

an investigation, are potentially influenced by psychological biases. Outside of healthcare, it is suggested that the impact of bias is broad, effecting what information is collected and how the analysis is carried out, by and with whom.[198] Bias has, in theory, the potential to cause an inappropriate focus on individual culpability, and narrow or skew the exploration and understanding of an incident's causation.[199] The impact of cognitive biases is further complicated by commonly held fallacies about their nature, for instance, that they only affect corrupt, malicious or incompetent individuals, that experts are 'immune', and that simply being aware of biases allows individuals to overcome their affect.[198] Outcome bias has not been explored, empirically, in the context of healthcare investigations and recommendation generation.

Exploring the literature on recommendation generation both confirmed that the observations from a local hospital perspective were in fact similar to those from hospitals across healthcare systems worldwide, and also highlighted gaps in the literature which could inform the aims and objectives of the thesis. It is these aims and objectives that will be presented in the next section.

2.4 Thesis aims & objectives

Chapter 1 presented some important background context, theories and concepts in which this PhD is situated. This chapter first discussed how this PhD was conceived as well as the approach to research, followed by initial inquiry (local hospital and literature) that informed the aims and objectives presented below. This thesis aimed to explore how to produce effective and achievable safety strategies from adverse event investigations in healthcare.

The three questions presented at the end of section 2.3.1 remain:

1) How are recommendations generated? What tools are used?
2) What makes a 'good' recommendation and how is this defined?
3) Do recommendations for improvements focus on individuals, and why?

The following specific objectives were developed in order answer these questions:

1) To explore the extant empirical knowledge about how hospitals:
    a. approach incident investigation before recommendation generation,
    b. how they generate recommendations,
    c. what recommendations are generated, and
    d. and what criteria are used to assess the quality or strength of recommendations made?
2) To examine the impact of outcome bias on the judgements and recommendations following hospital-based incident investigations.
3) To explore whether these biases might be reduced or eliminated through training or expertise in patient safety.
4) To identify criteria that would indicate a good investigation or recommendation and
5) see if it is possible to achieve consensus on what 'good' looks like in investigation and recommendation generation.

2.5 Methods

This section introduces the three research studies that were conducted to meet the objectives of this thesis, with more detailed methods discussed within each study. The approximate timeline of the studies throughout this thesis are illustrated in figure 2.3 below(page 62). The purple dotted arrows illustrate how the studies are connected, and more details are provided in figure 11.

2.5.1 Objective 1 was to explore the processes of recommendation generation in hospitals.

Munn and colleagues proposed the purpose of scoping reviews to be:[200]

- To identify the types of available evidence in a given field
- To clarify key concepts/ definitions in the literature
- To examine how research is conducted on a certain topic or field
- To identify key characteristics or factors related to a concept
- As a precursor to a systematic review
- To identify and analyse knowledge gaps

While a systematic review may be more valid when considering a specific clinical question such as feasibility or effectiveness of a treatment,[201,202] a scoping review may be more valid when aiming to identify or explore characteristics or concepts, as is this case for objective 1.[200] A scoping review would allow the examination of the extant literature on the process for generating recommendations, types of recommendations, and ways in which hospitals might judge recommendation effectiveness. An alternative methodological approach would have been to conduct a study within a single organisation, such as content analysis of investigation reports or interview study with investigators; both of which may have results in less generalisable findings. A brief summary of the scoping review methods is provided below, and the published study can be found on page 64.

Study 1: Exploring the "Black Box" of Recommendation Generation in Local Health Care Incident Investigations: A Scoping Review

Aim: The aim of this review was to examine the routine investigation and recommendation generation processes that occur in hospitals, considering the following questions:

1) What approaches to incident investigation are used before the generation of recommendations?
2) What are the processes for generating recommendations after a patient safety incident investigation?
3) What are the number and types of recommendations proposed?
4) What criteria are used, by hospitals or study authors, to assess the quality or strength of recommendations made?

Methods: Scoping Review, content analysis, categorization of extracted recommendations according the to AH.

Sources: peer-reviewed studies

2.5.2 Objectives 2 and 3 were to examine the impact of outcome bias on the judgements and recommendations following hospital-based incident investigations, and the impact of expertise on this. Vignette studies have long been used in psychology, sociology and healthcare research to examine people's attitudes and judgements towards scenarios.[192,194,195,203-208] Vignette studies allow researchers to present participants with carefully constructed scenarios in order to measure attitudes and opinions, and how these might change based on various manipulated variables.[206,209] There are a number of key advantages of vignette research. Firstly, vignette studies allow for the collection of large data sets in a relatively short time frame, and with relatively less expense than alternatives such as observational studies.[206,210] This is particular advantageous for responding to quantitatively focused research questions and subsequent statistical analysis.[210] Secondly, vignette studies allow for the control and manipulation of important variables, which would not be possible in an observational approach.[206] For these reasons vignettes were used to answer objective 2; the term scenario study was used and a brief summary of the methods is provided below, while the published study can be found on page 89.

Study 2: Investigators are human too: outcome bias and perceptions of individual culpability in patient safety incident investigations

Aim: The purpose of this study was to examine the impact of outcome bias on the judgements and recommendations following hospital-based incident investigations. We also explored whether these biases might be reduced or eliminated through training or expertise in patient safety. The following hypotheses were tested:

Hypothesis 1: increasing outcome severity is associated with increased judgements of responsibility and avoidability.

Hypothesis 2: increasing outcome severity is associated with increased judgements of the importance of investigation.

Hypothesis 3: increasing outcome severity is associated with more recommendations, and more punitive recommendations.

Hypothesis 4: expertise in patient safety will reduce outcome bias.

Methods: Experimental design/ scenario study, ordinal logistic regression models and multilevel linear models were used in the statistical analysis.

Sources: Questionnaires completed by participants recruited from three groups: (1) public, (2) healthcare staff, (3) people with expertise in patient safety or investigation of patient safety incidents.

2.5.3 Objectives 4 and 5 were to establish what 'good looks like' in terms of investigations and recommendations.

Before explaining the methodological approach to this question, it is important to explain why this aim related to investigations and recommendations, rather than just recommendations, which is the subject of this thesis. There are two main reasons:

1) Recommendations are generated as a result of the findings of an investigation. Lundberg and colleagues described a basic assumption underlying accident investigation, that "analysis of specific events will reveal patterns of underlying causes and conditions that if addressed by the right remedial actions can

prevent further events".[9] It is therefore reasonable to suggest that the methods or quality of the investigation might directly impact the quality of recommendations generated.[211] While studies such as that undertaken by Isherwood sought to examine the impact of different methodologies on generated recommendations,[153] the relationship between investigation and recommendation quality in healthcare investigations has not been empirically tested.

2) With reason 1) in mind, and in the interests of efficiency and to avoid wasting participants time with two separate studies, it seemed reasonable to identify criteria for both investigation and recommendation quality.

The lack of agreement about what constitutes a good recommendation was highlighted in section 2.3.2.2, along with the lack of practical guidance about how recommendations should be generated. Much of the guidance available for investigation was based on techniques such as root cause analysis, that have been transferred from other industries without sufficient attention to the differences of the healthcare context.[12,212,213] Despite the widespread adoption of incident investigation in healthcare,[157] few studies have critically examined the process.[10,178] Without a significant body of empirical research guiding how to judge healthcare investigation and recommendation quality, the Delphi technique had potential to move this area forward. The Delphi technique can be described as a series of questionnaires, with intervening feedback, shared with a group of experts, in order to gain consensus of opinion.[214] Stewart articulated the Delphi techniques as having the "capacity to capture those areas of collective knowledge that are held within professions but not often verbalized".[215] Delphi can be useful when there is a potential for differences of opinion and experience.[216] In the context of this study, for example, patients' perspectives and observations of safety can differ from those staff involved in patient safety,[217,218] and so it was reasonable to hypothesise that patients' opinion of what good looks like in investigation and recommendation generation may differ from others within an expert panel.

The Delphi technique supports the involvement of a heterogenous expert panel, and the potential to avoid domination of one opinion over another,[219] for instance investigators

over patients. A Delphi study had the potential to capture the experience and knowledge of a panel of experts, namely investigators, human factors experts, patient safety academics and those patients involved in incidents and investigations. Previous studies employing the Delphi technique successfully combined sometimes scant academic literature, with grey literature and expert opinion to achieve agreement on a range of topics.[220-222] It is for these reasons that a Delphi technique was considered appropriate for meeting objectives 4 and 5.

A brief summary of the methods is provided below, while the published study can be found on page 110.

Study 3: What does 'good' look like in healthcare incident investigations and recommendations? A modified Delphi study.

Aim: The aim of this study was to attempt to achieve consensus on what 'good' looks like in investigation and recommendation generation. More specifically aiming to identify criteria that would indicate a good investigation or recommendation.

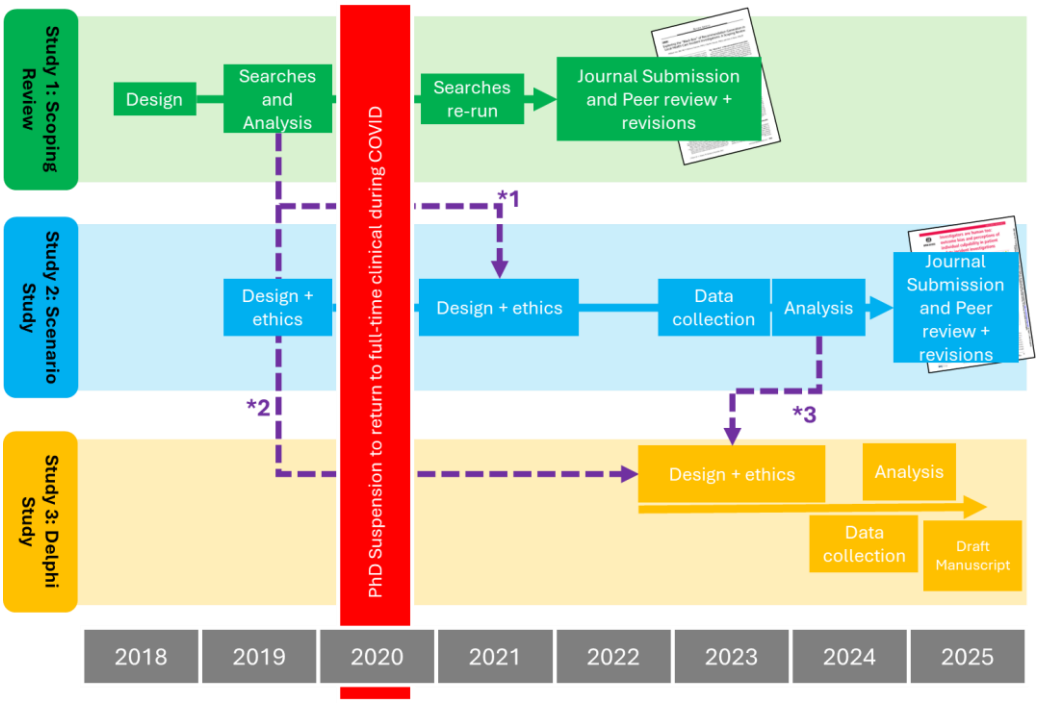
Methods: Delphi. Guided by an expert steering group, a 2-stage consensus-building approach was used: Stage 1) identifying and drafting candidate criteria that would indicate a good investigation or recommendation; Stage 2) conducting a 3-round modified Delphi process to identify, refine wording, and gain consensus on candidate criteria.

Sources(literature in stage 1 and participants in stage 2): In Stage 1, candidate criteria were drafted based on a review of peer-reviewed and grey literature and with the support of an expert steering group. In Stage 2, an expert panel was formed and presented with three rounds of questionnaires in order to review candidate criteria. Panel members had one or more of the following expertise:

- Public experience of investigations or first victims[223] (having been the victim or close relative/person with caring role for the victim of a patient safety incident and subsequently involved in the investigation)
- Practical experience of carrying out investigations (defined as having completed >10 investigations)
- Human factors and safety expertise (defined as having held an academic or practical job role in these areas for >12 months)
- Experience of designing or implementing interventions and system change, following investigations (defined as having held an academic or practical job role in these areas for >12 months, which may include quality improvement)*
- Responsibility for investigation processes or policy within local organisations or nationally (defined as having held a relevant job role for >12 months)

(*Examples might include patient safety leads, quality improvement leads, senior clinicians in hospitals)

Figure 2.3 Timeline and connection between thesis aims and studies. *The PhD was completed part-time, alongside clinical role as an acute medicine doctor and educator at the Hull York Medical School.*



**1 The scoping review provided further evidence for the continued tendency for "weaker" recommendations that focus on improving individuals' behaviour and practice, rather than the wider system deficiencies that contribute to incidents. This provided further evidence for the need to explore the reasons for this including the potential impact of outcome bias – explored in study 2. The analysis of recommendation and use of the Action Hierarchy in study one was also used to inform the design of the scenarios within study 2.*

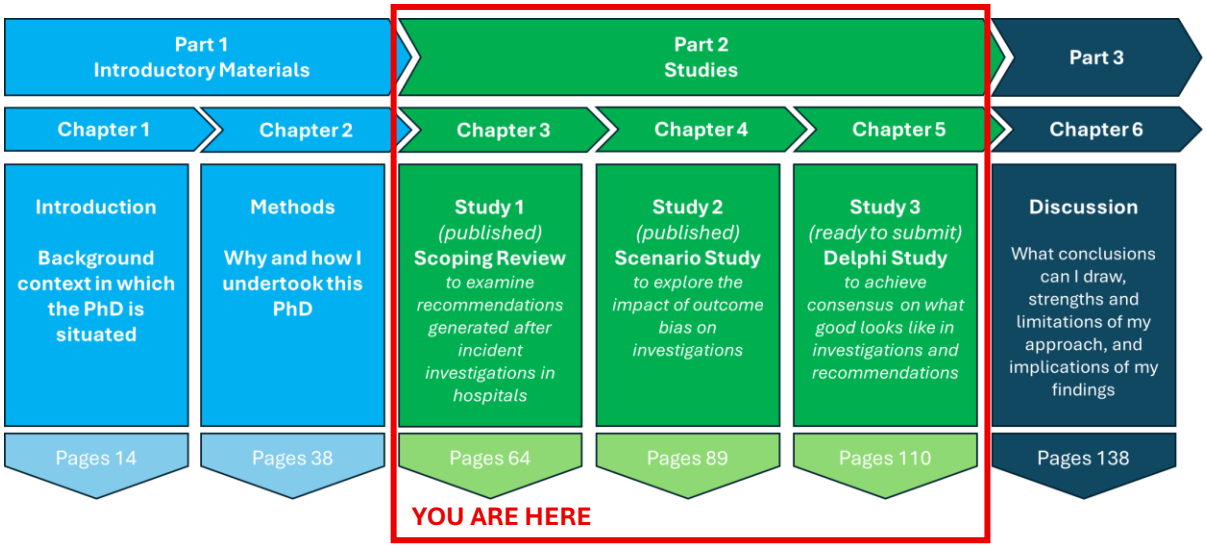
**2 The scoping review demonstrated confusion and a lack of consensus about the definition of recommendations and how to judge their effectiveness or quality. This prompted the need to attempt to establish criteria against which recommendations could be judged for effectiveness. With the presumed link between investigation quality and recommendation quality, I felt it would be sensible to design study 3 to consider both.*

**3 The scenario study demonstrated that when patient outcomes were worse following a patient safety incident, greater responsibility was assigned to healthcare staff involved, there was a stronger motivation to investigate, and an increased likelihood of punitive recommendations. Expertise in patient safety seemed to reduce, but not eliminate, these biases. These findings provided important information for the drafting of criteria for the modified Delphi (study 3), in that the management of bias is likely to be an important consideration in good quality investigations and recommendations.*

Chapter 2 Summary

Significant gaps in the research exist around how recommendations are generated within hospitals and how they should be judged for quality and effectiveness. The aims and objectives of this PhD and thesis, informed by the PhD candidates' front line patient safety experience and literature, are to address these gaps.

Part 2 – Studies



Chapters 3, 4 and 5 will present the studies within this thesis:

Chapter 3:(Study 1) Exploring the "Black Box" of Recommendation Generation in Local Health Care Incident Investigations: A Scoping Review

Published: Lea W, Lawton R, Vincent C, O'Hara, J. Exploring the "Black Box" of Recommendation Generation in Local Health Care Incident Investigations: A Scoping Review. Journal of Patient Safety 19(8):p 553-563, December 2023.

Chapter 4: (Study 2) Investigators are human too: outcome bias and perceptions of individual culpability in patient safety incident investigations

Published: Lea W, Budworth L, O'Hara J, Vincent C, Lawton R. Investigators are human too: outcome bias and perceptions of individual culpability in patient safety incident investigations *BMJ Quality & Safety* Published Online First: 10 February 2025.

Chapter 5: (Study 3) What does good look like in healthcare incident investigations and recommendations? A modified Delphi study.

Drafted for submission but not yet submitted.

Chapter 3:(Study 1) Exploring the "Black Box" of Recommendation Generation in Local Health Care Incident Investigations: A Scoping Review

Published: Lea W, Lawton R, Vincent C, O'Hara, J. Exploring the "Black Box" of Recommendation Generation in Local Health Care Incident Investigations: A Scoping Review. Journal of Patient Safety 19(8):p 553-563, December 2023.

Available online at:
https://journals.lww.com/journalpatientsafety/fulltext/2023/12000/exploring_the__black_box__of_recommendation.8.aspx#

3.1 Abstract

Background: Incident investigation remains a cornerstone of patient safety management and improvement, with recommendations meant to drive action and improvement. There is little empirical evidence about how—in real-world hospital settings—recommendations are generated or judged for effectiveness.

Objectives: Our research questions, concerning internal hospital investigations, were as follows: (1) What approaches to incident investigation are used before the generation of recommendations? (2) What are the processes for generating recommendations after a patient safety incident investigation? (3) What are the number and types of recommendations proposed? (4) What criteria are used, by hospitals or study authors, to assess the quality or strength of recommendations made?

Methods: Following PRISMA-ScR guidelines, we conducted a scoping review. Studies were included if they reported data from investigations undertaken and recommendations generated within hospitals. Review questions were answered with content analysis, and extracted recommendations were categorized and counted.

Results: Eleven studies met the inclusion criteria. Root cause analysis was the dominant investigation approach, but methods for recommendation generation were unclear. A total of 4579 recommendations were extracted, largely focusing on individuals' behavior rather than addressing deficiencies in systems (<7% classified as strong). Included studies reported recommendation effectiveness as judged against

predefined “action” hierarchies or by incident recurrence, which was not comprehensively reported.

Conclusions: Despite the ubiquity of incident investigation, there is a surprising lack of evidence concerning how recommendation generation is or should be undertaken. Little evidence is presented to show that investigations or recommendations result in improved care quality or safety. We contend that, although incident investigations remain foundational to patient safety, more enquiry is needed about how this important work is actually achieved and whether it can contribute to improving quality of care.

3.2 The "Black Box" of Recommendation Generation

Since the inception of the patient safety "movement," efforts to improve patient safety within hospitals have relied heavily on the retrospective investigation of adverse events.[3] Retrospective incident investigations as a mechanism for safety improvement are founded on an interpretation of safety theory, which proposes that errors are multifactorial in nature and that identifying and addressing organizational latent failures through investigation and recommendations will reduce future recurrence.[87,107]

In recent years, the generation of recommendations, after incident investigations, has come under increasing academic scrutiny.[10-12,15] This interest has occurred in parallel with the establishment of national-level independent investigatory bodies (e.g., HSIB in the UK, Norwegian Healthcare Investigation Board in Norway),[16,17] and in the UK, an ever increasing number of public inquiries and the ever expanding set of associated recommendations (e.g., Kirkup,[18] Ockenden,[224] Infected Blood Inquiries[225]). Therefore, exploring the act of recommendation generation is of increasing relevance as the number of recommendations across both local and national level investigation activity grows exponentially.

Although there are a plethora of aims and processes for investigations, a consistent feature is the production of recommendations. Despite 3 decades of incident investigation activity in health care,[157] few studies have critically examined the process.[10,178] In addition to the lack of empirical work examining recommendation generation, there is a lack of practical guidance, on the generation of recommendations.[15] One systematic review used a modified version of the National Institute for Occupational Safety and Health hierarchy of risk controls to categorize the recommendations from included studies,[10,184] concluding that 80% of recommendations were "weak," that is, unlikely to result in significant improvements in safety or risk reduction. Furthermore, Hibbert and colleagues undertook a retrospective study,[100] following investigations within an Australian regional health system. The study used and modified the U.S. Department of Veteran Affairs action hierarchy (AH) to

categorize recommendations as strong, medium, or weak and concluded that only a small number of recommendations were strong and the most common types of recommendations involved reviewing or enhancing policies/guidelines/documentation as well as training and education.[100] It is important to note that these issues extend beyond health care. Indeed evidence suggests that a lack of guidance and a plethora of other sociotechnical factors impede the generation, implementation, and evaluation of recommendations across safety investigations in contexts such as rail, maritime, and nuclear.[9,15]

3.3 Recommendation Generation Within Local Health Care Investigations

Despite the centrality of incident investigation and recommendation generation within patient safety policy globally, there is a surprising lack of understanding about what actually happens in local health care settings with respect to this important activity. In particular, there is a lack of empirical focus and consensus about recommendation generation by people conducting investigations at a local health care organization level.[12,157] This review therefore aims to examine the extant empirical knowledge about this issue. We have focused on hospital settings rather than primary/community care because of the fundamentally different ways in which care is delivered and case mix,[226] as well as the relatively lower level of incident reporting and relevant published literature in primary care.[36,226,227]

3.4 Scoping Review Aims

The purpose of this review was to consider the following questions:

1. What approaches to incident investigation are used before the generation of recommendations?
2. What are the processes for generating recommendations after a patient safety incident investigation?
3. What are the number and types of recommendations proposed?
4. What criteria are used, by hospitals or study authors, to assess the quality or strength of recommendations made?

3.5 Methods

We conducted a scoping review, following the preferred reporting items for systematic reviews and meta-analyses extension for scoping reviews guidance.[228]

3.5.1 Sources and Searches

Searches were performed on February 28, 2019, and January 30, 2021, using MEDLINE, EMBASE, PsychINFO, and CINAHL. Search terms were iteratively developed to capture the key phases of incident investigation including terms for the incident, investigation, and subsequent recommendations (see appendix 2 for search terms, page 197). Searches were restricted to English language and studies published since 1999, when the Institute of Medicines' seminal report, To Err Is Human, was published,[2] prompting greater focus on patient safety.

3.5.2 Study Selection

The aim of this review was to examine the routine investigation and recommendation generation processes that occur in hospitals.

Studies were included if they reported on a series of incidents occurring in the hospital, which were chosen for investigation by hospital-based staff, who also generated subsequent recommendations. Studies reporting on incidents from any clinical context or level of harm were included.

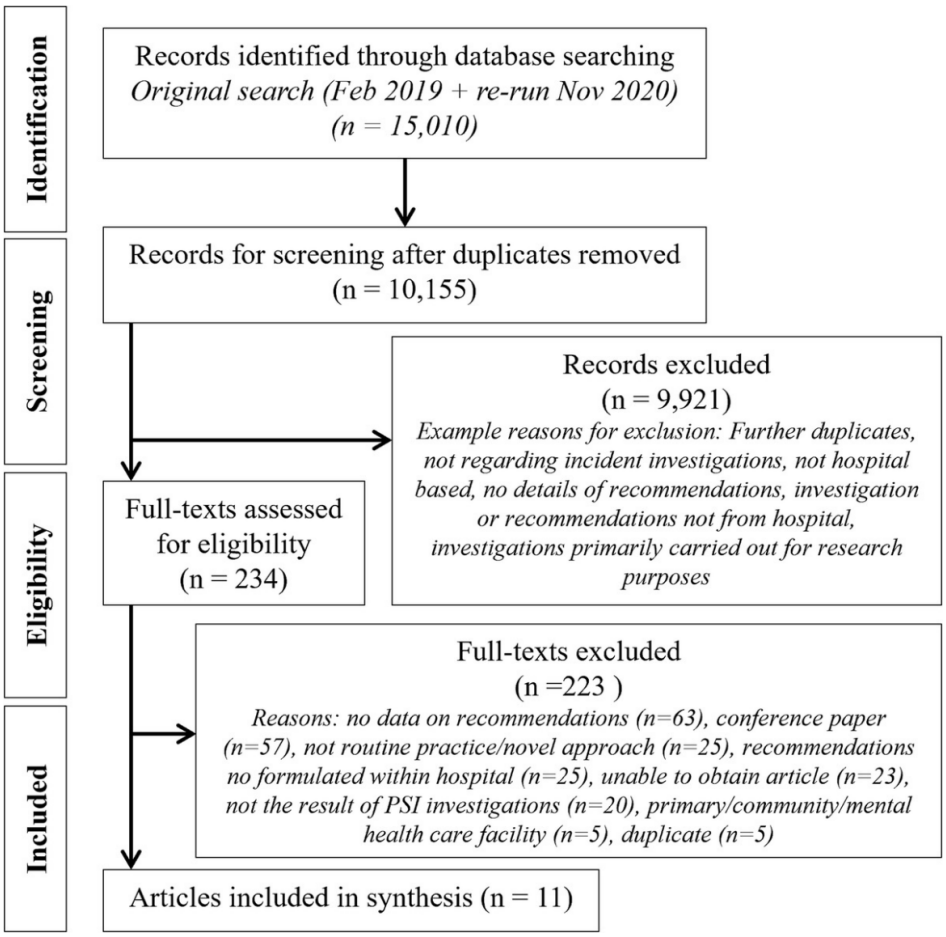
Studies were excluded if they reported data from the following:

1. Community, primary care, or primarily mental health care
2. Investigations/recommendations carried out or proposed outside of a hospital, for instance, by an external research team or regional organization
3. Investigations primarily carried out for the purposes of research
4. Not published/peer-reviewed (e.g., conference papers)

Searches yielded 15,010 articles. The article title and abstracts were reviewed by W.L. Random samples of 5% (n = 720) were screened independently by both J.O.H. and R.L. to check congruence. A total of 246 articles were selected for full-text review. Full-text

screening was undertaken by W.L., with 10% independently screened by each of J.O.H. and R.L. (n = 20). Any discrepancies were discussed and resolved between authors. Eleven articles met the inclusion/exclusion criteria (all agreed with W.L., J.O.H., and R.L.) and contributed to the review (Fig. 3.1). Regular meetings with the other author (C.V.) allowed discussion of article eligibility.

FIGURE 3.1: PRISMA-ScR flow diagram for study selection.



3.5.3 Data Extraction and Quality Assessment

The purpose of the review was to examine the nature of recommendations proposed within hospitals, which was not the primary aim of all the included studies, but those included did contain empirical data on recommendations.

We assessed study quality using the Quality Assessment for Diverse Studies (QuADS) tool.[229] This tool is a well-cited approach to assessing the quality of methodologically heterogeneous studies, which demonstrates reliability and validity.[229,230] After discussion of the application of the tool and relevance of quality scoring by all the authors, W.L. reviewed and scored all included articles. A random sample (n = 4 [36%]) of studies were independently reviewed and scored by J.O.H. and R.L., with disagreements resolved with discussion.

3.5.4 Data Synthesis and Analysis

To address research questions 1, 2, and 4, we undertook content analysis of the included studies using 4 stages; decontextualization, recontextualization, categorization, and compilation.[231] First, authors read and made themselves familiar with the included studies before extracting "meaning units" of text relevant to answering the aims of the review (decontextualization). After extraction of meaning units, the remaining article text was checked for further relevant content (recontextualization). Next the extracted meaning units were split into specific areas relevant to each research question; the word count was reduced without losing the meaning/content (categorization). The research questions were answered by condensing the extracted text using the original study terms and language, as well as providing numerical counts of how often content was reported across the studies.

To address research question 3, we used the AH, proposed by the U.S. Veteran Affairs National Center for Patient Safety to categorize the recommendations extracted from included studies.[97,100,101,182,183] Recommendations from the included studies were discussed by all the authors across 2 meetings and assigned to the core categories of the AH, then counted, to report frequency. If, after discussion, it was felt that a recommendation or category of recommendations did not fit into one of the AH categories, a new category was created and agreed.

3.6 Results

The characteristics of included studies (n = 11) are summarized in Table 3.1. Included studies contained 4680 recommendations from 2818 investigations carried out across 171 hospitals.

3.6.1 Country of Origin

Included studies were conducted in the United States (n = 4), the United Kingdom (n = 2), and Australia (n = 2), with one each from the Netherlands, Brazil, and Hong Kong.

3.6.2 Clinical Context and Incident Harm

Studies reported data from across all clinical specialties (n = 6), pharmacy/medication (n = 1), anesthesia and intensive care (n = 2), and pediatric care (n = 2). Incidents reported within studies varied in their type (e.g., delay in care, fall, dispensing of medication) and resulting harm (see Table 1 for more detail).

3.6.3 Quality Assessment

The included studies demonstrated an average QuADS score of 56% (range, 26%–69%) Five of 11 studies lacked theoretical underpinning such as the discussion of an accident causation model. Half of the studies did not report, in sufficient detail, the justification of sampling or selection of data collection tools. Six studies had no evidence that research stakeholders had been involved in their planning or conduct. Four studies had limited or no discussion of their strengths or limitations. No studies were excluded based on quality.

Table 3.1 Included studies summary table

Author (year)	Reference no	Country	Clinical context	No of hospitals	No. of incidents analysed	Types of incident(s)/incident(s) relating to	Levels of harm of investigated incident(s)	No. of recommendatio	How are recommendations categorised?
Corwin (2017)	95	United States of America	Intensive care	47	70	Delay in care; medication; medical procedure; equipment failure; removal of lines, catheters, tubes, drains; transfusion, elopement; discharge; suicide attempt; fall; airway/ventilation	Incidents causing harm only	276	Categories developed by authors based on included investigations recommendations
Figueiredo (2018)	96	Brazil	Tertiary general hospital	1	1316	Drug supply chain; Fall; Pressure ulcer ; Other skin lesions; Surgical procedure (relation with laterality); Transfusion process; Unplanned withdrawal of catheter, drain, tube or catheter; Identification of patient; Loss of sample; Bruise; Extravasation; Delay in exam/procedure completion; Prolonged fasting; Failure to release the technical report; Evasion;	No harm and harm incidents included	1326	Categories developed by authors based on included investigations recommendations

						Technical or equipment/material handling failure; Failure to identify material/instruments; Related to childbirth; Nutritional therapy; Healthcare related infection; Failure during technique, procedure or transportation; Death and Others			
Hamilton (2019)	97	Australia	Hospitals / Paediatric	16	42	Delayed diagnosis; delayed recognition or response to a deteriorating patient; procedural adverse event; patient identification or procedure mismatching; medication adverse event; delayed definitive treatment; unexpected death/event following hospital presentation or admission; testicular torsion delayed diagnosis or management	Incidents causing harm only	150	US Department of Veteran Affairs' criteria, or 'action hierarchy' (VA Action Hierarchy)
Hibbert (2018)	100	Australia	Hospitals in region	36	227	Clinical process/procedure; falls; behaviours; problems with diagnosis; problems with procedures or interventions; wrong patient/body part; inpatient suicide; retained instruments/other; gas	Incidents causing harm only	1137	US Department of Veteran Affairs' criteria, or 'action hierarchy' (VA Action Hierarchy)

						embolism; ABO incompatible blood transfusion; medication error; maternal death; wrong infant discharged			
Irwin (2011)	93	United Kingdom	Hospitals in region / Pharmacy	23	573	Dispensing of medications	No harm and harm incidents included	251	Categories developed by authors based on included investigations recommendations
Kellogg (2017)	99	United States of America	Academic Medical Centre	1	302	Procedure complication; cardiopulmonary arrest; neurological deficit; retained foreign body; pulmonary/arterial embolus; birth complication; medication administration error; incorrect procedure/study; sepsis; wrong-site surgery/procedure; devastating illness; myocardial infarction; haemorrhage/haematoma; arrhythmia; unknown cause of death; adverse medication event; compartment syndrome; fall, inpatient;	No harm and harm incidents included	499	Categories developed by authors based on included investigations recommendations

						event proximate to discharge; self-harm; electrolyte disturbance; assault, inpatient; bowel perforation; equipment failure; sleep apnoea; ventilation complication			
Kwok (2020)	101	Hong Kong	Hospitals in region	43	214	Surgery/interventional procedure involving the wrong patient or body part; retained instruments or other material after surgery/interventional procedure; ABO incompatibility blood transfusion; intravascular gas embolism resulting in death or neurological damage; death of an inpatient from suicide; maternal death or serious morbidity associated with labour or delivery; infant discharged to wrong family or infant abduction; other adverse events resulting in permanent loss of function or death; medication error which could have lead to	No harm and harm incidents included	760	US Department of Veteran Affairs' criteria, or 'action hierarchy' (VA Action Hierarchy)

						death or permanent harm; patient misidentification which could have led to death or permanent harm			
Morse (2012)	182	United States of America	Paediatric hospital	1	20	Medication event; delayed identification of clinical deterioration; equipment failure; enteral feeding via central line; breast milk event; unsterile surgical equipment; inappropriate patient behaviour in the "play room"; significant tissue injury; name change patient identification; readmission event; wrong site MRI under general anaesthesia	No harm and harm incidents included	78	US Department of Veteran Affairs' criteria, or 'action hierarchy' (VA Action Hierarchy)
Robbins (2020)	232	United Kingdom	University Hospital	1	22	No details	Incidents causing harm only	101*	Hierarchy of intervention effectiveness (people versus system focused) (Cafazzo 2012)
Van der Starr (2014)	233	The Netherlands	Neonatal / Paediatric intensive care	1	17	Medication errors; procedural; unanticipated death, unanticipated resuscitation; nursing care	No harm and harm incidents included	84	Factors influencing clinical practice devised by Vincent et al (2000)

Zeng (2016)	94	United States of America	Hospital	1	15	Reported most common: communication failure between team members; tasks to automate for dosimetry; simulation issues; IT system failures; scheduling issues.	No harm and harm incidents included	18	Categories developed by authors based on included investigations recommendations
TOTAL				**171**	**2818**			**4579**	
*lacked detail to enable categorisation, therefore not included in table 3									

3.6.4 RQ1) Approaches to Incident Investigation Used Before the Generation of Recommendations

Nine studies reported using root cause analysis (RCA),[94,95,97,99,100,101,182,232,233] 3 used both RCA and the London Protocol,[94,100,233] and the remaining 2 used no specific tool or method.[93,96] Four studies reported that a team of 2 to 8 staff (physicians, nurses, and managers) undertook the investigation,[94,99,101,233] and 2 reported specific investigator training.[232,233] The remaining studies did not provide these details.

As part of the investigation process, 3 studies reported interviewing staff,[94,99,233] one of which specified that incidents were reconstructed from a median of 6 interviews (n = 3–15).[233] One study reported that parents of children involved in incidents were interviewed "if felt to be useful," and this occurred in 2 of 17 incidents.[233]

Four studies reported on the time spent undertaking investigations. This was highly variable, ranging from 3 to 90 hours.[94,182,232,233] Three studies reported that investigations should be completed within a set period of time, ranging from 30 to 60 days,[95,99,101] although they did not specify if this was from when the incident occurred or was reported, or the decision to investigate was made.

3.6.5 RQ2) The Processes for Generating Recommendations After A Patient Safety Incident Investigation

None of the included studies reported using specific tools or methods for recommendation generation. One article reported that staff and parents were invited to suggest recommendations, whereas none of the remainder reported this kind of stakeholder involvement.[233] Eight studies proposed that recommendations should prevent incident recurrence[94,95,97,99-101,232,233] and eliminate, mitigate, or reduce a risk, hazard, or "root causes."[95,99,101,232] No purpose or aim for recommendations was stated in the remaining 3 studies.

3.6.6 RQ3) The Number and Types of Recommendations Proposed

A variety of terms were used to describe the recommendations generated after investigations. We present these terms in Table 3.2, but because the terms were not clearly defined within the studies, we were not able to determine differences or similarities and have therefore reported them as written. A total of 4579 recommendations were extracted from 10 included studies (Table 3.3), with an average of 3.7 (1–5) per investigation. Recommendations were not extracted from the 11th included study because of insufficient detail to enable categorization.[232] Six studies assigned recommendations to predetermined categories based on (i) the U.S. Department of Veteran Affairs' criteria or AH,[97,100,101,182] (ii) factors influencing clinical practice devised by Woloshynowych et al,[107,233] or (iii) the "hierarchy of intervention effectiveness" (people versus system focused).[232] The remaining 5 studies developed their own categories based on analysis of their included recommendations.[93-96,99]

Table 3.2 Terms used to describe the recommendations following investigations

	Frequency	Study reference
Recommendations	5	Hibbert(2018), Hamilton(2019), Kwok(2020), Corwin(2017), Van der Starre(2014)
Action(s)	5	Kwok(2020), Corwin(2017), Figueiredo(2018), Kellogg(2017), Robbins(2020)
Action plan(s)	2	Morse(2012), Zeng(2016)
Corrective actions/action plans	2	Morse(2012), Robbins(2020)
Solutions	2	Kellogg(2017), Robbins(2020)
Process improvements	1	Zeng(2016)
Interventions	1	Irwin(2011)
Risk reduction strategies/measures	1	Morse(2012)
Preventative measures	1	Van der Starre(2014)
Recommended actions	1	Corwin(2017)
Managerial responses	1	Irwin(2011)
Error management strategies	1	Irwin(2011)
Risk controls	1	Van der Starre(2014)
Process improvement projects	1	Zeng(2016)

Education or training represented the most common recommendation (27.2% [n = 1257]), followed by new procedure/memorandum/policy (15% [n = 676]), change of

process or routine (10.7% [n = 500]), and adjustment/improvement to policy or guideline (6.7% [n = 306]). Fourteen percent of the extracted recommendations were too vague or unclear to categorize. Table 3.3 shows the full breakdown of recommendations by category. Recommendation categories 1 to 26, in Table 3.3, are from the AH,[97,100,101,182] and categories 27 to 36 are those proposed by the study authors. Six hundred fifty-six recommendations were categorized as "vague/unclear" either by the authors of the included studies or authors of this review during analysis. Examples of vague/unclear recommendations included "Medication incident action plan implemented" (n = 3)(Irwin 2011); "policy, procedure and process actions" (n = 5);[95] and "provide counseling" (n = 280).[96]

Table 3.3 Recommendations extracted from included studies

AH strength of recommendations	Recommendation category number	Recommendation Categories	n	% within Action Hierarchy	% all recommendations
Strong	1	Standardize on equipment or process	66	2.0%	1.4%
	2	Architectural/physical plant changes	60	1.8%	1.3%
	3	Tangible involvement by leadership	44	1.3%	1.0%
	4	New devices with usability testing	23	0.7%	0.5%
	5	Engineering control (forcing function)	16	0.5%	0.3%
	6	Simplify process	14	0.4%	0.3%
		Total Strong	**223**	**6.8%**	**4.9%**
Medium	7	Adjust or improve a policy or guideline	306	9.4%	6.7%
	8	Enhanced documentation or communication	170	5.2%	3.7%
	9	Audit undertaken	149	4.6%	3.3%
	10	Checklist or cognitive aids	90	2.8%	2.0%
	11	Software enhancements or modifications	69	2.1%	1.5%
	12	Analyse/inspect/review use or appropriateness of equipment	37	1.1%	0.8%
	13	Review rostering/appropriateness of staff mix	32	1.0%	0.7%
	14	Increase in staffing/decrease in workload	17	0.5%	0.4%
	15	Standardized communication tools	12	0.4%	0.3%
	16	Education using simulation-based training, with periodic refresher sessions and observation	13	0.4%	0.3%
	17	Redundancy	9	0.3%	0.2%
	18	Eliminate/reduce distractions	9	0.3%	0.2%

	19	New [clinical] team	5	0.2%	0.1%
	20	Eliminate look- and sound-alikes	1	0.0%	0.0%
		Total Medium	**919**	**28.2%**	**20.1%**
Weak	21	Training	1257	38.5%	27.5%
	22	New procedure/memorandum/policy	676	20.7%	14.8%
	23	Meeting to discuss event/staff made aware of event	105	3.2%	2.3%
	24	Staff asked to provide written reflective statement or staff informed/notified/warned	30	0.9%	0.7%
	25	Double checks	27	0.8%	0.6%
	26	Warnings	25	0.8%	0.5%
		Total Weak	**2120**	**65.0%**	**46.3%**
		Total categorised within AH	**3262**	**100.0%**	
new categories	27	Vague/unclear	656		14.3%
	28	Change of process/routine	500		10.9%
	29	Additional study/analysis	121		2.6%
	30	Risk assessment/management/risk register	12		0.3%
	31	Supervision	12		0.3%
	32	Involvement of external organisation (external investigating or contacted as part of investigation)	10		0.2%
	33	New staff role	1		0.0%
	34	Purchase new equipment	3		0.1%
	36	Adjustments to team expertise/make-up	2		0.0%
		Total recommendations within new categories	**1317**		**28.8%**
		Total recommendations extracted	**4579**		**100.0%**

3.6.7 RQ4) Criteria Used to Assess the Quality or Strength of Recommendations Made

Two of 11 articles reported that the original internal hospital investigations made judgments of recommendation "quality" or "strength".[94,95] One study reported that the hospital prospectively tagged incidents to identify trends and therefore monitor for process improvements, although it did not report any data in relation to this.[94] Another study reported that implemented action (n = 277) effectiveness was rated by local managers as "much better" (47.4%), "better" (37.0%), "same"(7.4%), "worse" (0%), or not reported or measured (8.2%).[95] Although none the studies provided comprehensive data on incident recurrence, one study reported that similar incidents did reoccur despite multiple investigations.[99]

Included studies, in secondary analysis, used a range of terms or phrases to "judge" recommendations as follows.

- Effectiveness (Hibbert et al,[100] Kwok et al,[101] Corwin et al,[95] Figueiredo et al,[96] Kellogg et al,[99] van der Starre et al,[233] Robbins et al[232])
- Strength (Hibbert et al,[100] Morse and Pollack,[182] Hamilton et al,[97] Kwok et al,[101] Kellogg et al[99])
- Whether implemented (Morse and Pollack,[182] Hamilton et al,[97] Corwin et al,[95] Kellogg et al,[99] van der Starre et al[233])
- Aimed at system level improvements or modifying processes (Morse and Pollack,[182] Kwok et al,[101] Kellogg et al[99])
- Likelihood they would prevent incident recurrence (Morse and Pollack,[182] Kellogg et al,[99] van der Starre et al[233])
- Quality (Morse and Pollack,[182] Robbins et al[232])
- Sustainability (Hibbert et al,[100] Kellogg et al[99])
- Efficacy (Hamilton et al[97])
- Innovation (Robbins et al[232])
- Level of impact (Morse and Pollack[182])

In 8 of 11 studies, authors discussed their approach to judging recommendations.[95,97,99,100,101,182,232,233] Four studies judged recommendations as strong, intermediate, or weak based on the AH with some variations in the category descriptions and/or addition of further categories.[97,100,101,182] One study referenced a "Model of Sustainability and Effectiveness in RCA Solutions,"[99,234] WHEREAS another reported effectiveness of recommendations according to the "Hierarchy of Intervention Effectiveness," which proposes that "system-focused changes have greater impact".[232,235] One article commented on recommendation likelihood of preventing incident recurrence,[233] based on a classification of recommendation strength (weak, medium, strong) proposed by the New South Wales Root Cause Analysis Review Committee.[148]

3.7 Discussion

To the author's knowledge, this review represents the first review of the extant empirical evidence for the practice of generating recommendations in hospitals, specifically examining how and what recommendations were generated, as well as the way in which their effectiveness was judged. This process is central to the efforts to improve patient safety and health care quality globally. Our review highlights the paradoxical situation that, despite the ubiquity of recommendation generation, very little is known about it in practice. Our findings suggest that, although RCA dominates as the approach to investigation, there are no specific tools or approaches used to generate recommendations. Recommendations focus on training or adding or improving policies. In other words, recommendations largely focus on staff knowledge and skills. There is a lack of agreement in the literature on how effectiveness of recommendations should be judged, meaning that there is very little understanding of what makes a "good" recommendation. These findings raise some important issues, which we will address in turn.

3.7.1 Recommendation Generation Is Confused and Unclear

The variety of terms used to describe recommendations (Table 3.2) and lack of consensus for categorization suggests differences in vision and purpose at best, and confusion and disagreement at worst. Although this review provides some steer in terms of the espoused investigation techniques, the actual process of how investigation outcomes result in specific recommendations remains opaque. We found that, beyond the investigators, there are committees or teams within hospitals as well as within local or regional organizations that review investigations and their findings; although what role these groups had in selecting or modifying recommendations is unclear. Studies in the wider literature have attempted to explore this process in practice. Braithwaite et al found a number of challenges to RCA such as time constraints, lack of resources, and unwilling colleagues.[147] Another study suggested that recommendations may actually be related to other ongoing improvement work; that is, the incident was used to support existing agendas rather than to generate new findings.[174] Furthermore, an ethnography

of investigations identified attempts by investigators to manage scrutiny and maintain reputations, and concluded that a failure to appreciate the complex organizational agendas as well as social and political influences on recommendation generation would likely hamper improvements in patient safety.[159] Beyond health care, studies of investigations from other domains, such as nuclear and rail, have demonstrated that the design of approaches to investigation and associated manuals lack emphasis or detail on the generation and evaluation of recommendations.[15] Another cross-domain study identified that there are a large number of cognitive, political contextual factors that influence the investigation and recommendation generation process, such as cost-benefit analysis, willingness of stakeholders to engage, or the experience or knowledge base of the investigator.[9] Collectively, these studies suggest that the generation of recommendations is likely to be a highly complex sociopolitical process with many stages and influences.[12,147,156,159,174] New approaches and tools for recommendation generation[179,180,181,236] are more likely to be successful if adapted and designed relative to the unique and complex context of health care.[213,237] Further research to understand the reality of the movement from investigation to recommendation generation is therefore important.

3.7.2 Recommendations Are Classified as Weak and Lack System Focus

This review identified that less than 7% of the extracted recommendations might be considered “strong” or system-focused, such as standardizing equipment, architectural changes, or simplifying processes. Our findings provide further evidence for the continued tendency for “weaker” recommendations that focus on improving individuals’ behavior and practice, rather than the wider system deficiencies that contribute to incidents. This tendency, shown in numerous studies from across the globe,[10,238-242,243-245] suggests explanatory reasons beyond national culture or specific differences in health care systems and is completely at odds with health care policy and safety research.[107,148,183] Furthermore, it would suggest that, globally, health care organizations may have some way to go toward achieving a more just culture, with this focus on weaker individual-focused recommendations both reflecting this and serving to reinforce it.[87]

Root cause analysis and frameworks, used to support investigation, have themselves been identified as narrowing the view of causation[12] or giving greater attention to causative factors relating to individuals.[144] With a tendency for investigations to identify individual factors,[144] it is perhaps not surprising that recommendations are targeted at the same level. Other reasons for a lack of system-level recommendations include lack of investigator training, expertise,[10] or health care–tailored guidance,[107] and difficulty in designing and implementing at the system level.[184,237] This review highlights the continued predominance of RCA, despite the growing number of alternatives that might broaden investigations and identify a wider range of contributory factors and subsequent recommendations.[153,246-249]

3.7.3 It Is Not Clear How to Judge Recommendations

Although the focus of recommendations at the weaker individual level has been widely challenged, a further compounding problem with recommendation generation is the lack of agreement on how to judge their effectiveness and what makes a "good" recommendation. The range of terms, in our included studies, such as "strength," "quality," "sustainability," and "implementability," indicates the complex nature of judging recommendations. Our review found 2 broad approaches: (i) the use of predefined hierarchies of recommendation effectiveness and (ii) assessing the effectiveness of recommendations over time.

Starting with hierarchies, this review demonstrates not only their widespread use but also variety and variation.[10,100,101,148,182,183,234,235,250,251] These hierarchies, largely originating from non–health care settings,[237,252] are used in health care with minimal empirical evidence.[234] They generally propose that recommendations targeted at the individual level (e.g., training and reminders) are weaker than those at the system level (e.g., equipment design). Before this review, there have been challenges of the use of hierarchies to predict recommendation effectiveness,[236,237] with arguments that recommendations should be judged on how well they align with the identified risks and context,[181] their likelihood of effecting necessary change,[253] or level of system targeted for change.[236] Our review suggests that hierarchies may not yet be widely used in

practice, but with the growing number of variations and lack of consensus, they have the potential to cause confusion for hospital safety teams looking to adopt evidence-based approaches. Beyond the need for empirical evaluation of these options, we suggest that future research will also need to consider the practical application of these in health care.

The second approach to judging recommendation effectiveness seems to be "post-hoc" measures, more specifically assessing what difference is made to processes and outcomes, as well as future incident occurrence. In problem solving, determining the effectiveness of solutions is a key step.[252] There is a surprising absence of post-hoc measures reported within the included studies, with none of the included studies comprehensively reporting the rates of incidence recurrence. With "the prevention of incident recurrence" being the most commonly quoted reason for incident investigation, it is of note that these data are lacking within this review, as well as the wider literature.[12,107,156]

Beyond this review, numerous studies have indicated that incident reporting systems (key for incident identification) only detect a minority of incidents that actually occur, and this number may be even lower for incidents resulting in harm.[32,63,73] Incident recurrence may be a poor marker of investigation success, if reporting remains unreliable. We contend that more research is needed to consider specifically what measures are appropriate for measuring recommendation or investigation effectiveness.

Although Reasons' organizational accident model is central to much of health care investigation practice,[87,107] the included studies demonstrate a lack of translation of the complexity and nuance of the original model. For instance, the recommendations largely focus on reducing error rates rather than putting in place defences to more broadly improve system safety and quality or reduce the impact of an error if it does occur. The studies included within this review provide no evidence that carrying out investigations and generating recommendations improve the quality or safety of care. Furthermore, there seems to be little consideration of the potential negative consequences of recommendations themselves.

3.8 Limitations

Despite the volume of incident reporting and investigation within health care, there is a relative lack of peer-reviewed research with empirical data from "real-world" hospital investigations. Relevant studies may have been excluded if there was ambiguity as to whether they reported data from usual practice within hospitals, as this was the focus of the review. Because of the lack of studies exploring the specific aims of this review, the included study's aims were not necessarily aligned with the aims of the review, rather relevant empirical data were extracted. Many of the included studies do not report the entire investigation process in detail or the effect of recommendations, which has impacted our ability to answer some of the review questions. It was not possible to analyse recommendations at the incident level, which would have allowed us to identify the proportion of recommendations at the individual and system levels. We recognize that this would be an important area for future research. Because we have focused on internal hospital investigations, as opposed to those at a regional or national level, there is a chance that this is one reason there are less observed recommendations targeting those contributory factors or organizations external to the hospital; internal hospital investigations may be more likely to focus on what they perceive they can change.[9] This review has focused on the generation of recommendations, but no assumption is made that "good" recommendations will necessarily improve safety. Implementation of recommendations and the challenges and barriers is another important factor to consider but was beyond the scope of this review.

3.9 Conclusions

The aim of this review was to explore hospitals' approaches to incident investigation, recommendation generation, the types of recommendations proposed and how their effectiveness is judged. Although RCA dominates as the approach to investigation, how recommendations are selected remains unclear. Recommendations are generally classified as weak, focusing on improving individuals' skills, knowledge, and understanding so as to change behaviour rather than addressing deficiencies in the systems in which staff work. Our review demonstrates a lack of evidence and

consensus regarding how recommendations should be judged for effectiveness. We argue that greater clarity is needed in terms of the purpose of investigations and the language used to describe them. Furthermore, empirical work needs to explore and explicate how to generate appropriate recommendations, as well as how these approaches are adopted within the complex sociotechnical context of health care.

Finally, we suggest that, although incident investigations remain foundational to patient safety measurement and improvement, more enquiry is needed about their effectiveness or impact. The generation of recommendations themselves is only one step in the process. Both policy and practice will also need to engage with the growing body of literature and adopt a more evidenced-based approach to investigation and recommendation selection.

Chapter 4: (Study 2) Investigators are human too: outcome bias and perceptions of individual culpability in patient safety incident investigations

Published: Lea W, Budworth L, O'Hara J, Vincent C, Lawton R. Investigators are human too: outcome bias and perceptions of individual culpability in patient safety incident investigations *BMJ Quality & Safety* Published Online First: 10 February 2025.

Available online at: https://qualitysafety.bmj.com/content/early/2025/02/09/bmjqs-2024-017926

4.1 Abstract

Background: Healthcare patient safety investigations inappropriately focus on individual culpability and the target of recommendations is often on the behaviours of individuals, rather than addressing latent failures of the system. The aim of this study was to explore whether outcome bias might provide some explanation for this. Outcome bias occurs when the ultimate outcome of a past event is given excessive weight, in comparison to other information, when judging the preceding actions or decisions.

Methods: We conducted a survey in which participants were each presented with three incident scenarios, followed by the findings of an investigation. The scenarios remained the same, but the patient outcome was manipulated. Participants were recruited via social media and we examined three groups (general public, healthcare staff and experts) and those with previous incident involvement. Participants were asked about staff responsibility, avoidability, importance of investigating and to select up to five recommendations to prevent recurrence. Summary statistics and multilevel modelling were used to examine the association between patient outcome and the above measures.

Results: 212 participants completed the online survey. Worsening patient outcome was associated with increased judgements of staff responsibility for causing the incident as well as greater motivation to investigate. More participants selected punitive recommendations when patient outcome was worse. While avoidability did not appear to be associated with patient outcome, ratings were high suggesting participants always

considered incidents to be highly avoidable. Those with patient safety expertise demonstrated these associations but to a lesser extent, when compared with other participants. We discuss important comparisons between the participant groups as well as those with previous incident involvement, as victim or staff member.

Interpretation: Outcome bias has a significant impact on judgements following incidents and investigations and may contribute to the continued focus on individual culpability and individual focused recommendations observed following investigations.

4.2 Introduction

Globally, efforts to improve patient safety have relied heavily on the retrospective investigation of patient safety incidents.[3] This approach is founded on an interpretation of safety theory, which proposes that errors are multifactorial in nature and that identifying and addressing organisational latent failures through investigation and generating recommendations will reduce future recurrence.[87,107]

Recent publications[10,12,100,238-242,244] including our own review[254] have identified that the overwhelming majority of recommendations developed following incident investigations would be categorised as 'weak' according to the framework developed by Hibbert and colleagues.[100] Rather than addressing latent failures of the system (eg, design of equipment), the target of recommendations is most often on the behaviours of individuals, such as reminders, writing or rewriting policies and (re-)training staff. The patient safety movement has struggled to shift the focus from people to systems; and this may be a reason why we are still not 'learning' from patient safety investigations and therefore not reaping the benefits of careful analysis.[12,156,255] In fact, it is now acknowledged that investigations can, themselves, compound or add harm to those involved or affected by the incident, investigation or subsequent recommendations.[256] To address these problems, we need to better understand the flaws in the incident investigation process itself.

There has long been evidence that our judgements (attributions) of individual responsibility or culpability are driven by the outcome of an accident or adverse event.[190,191] In other words, the same behaviour (parking on a slope without putting the handbrake on) is judged more harshly when the outcome is bad (the car moves and runs someone over), than when the outcome is not (the car does not move or moves but no-one is harmed). While the original studies of this bias were conducted within the field of road traffic accidents or legal settings, subsequent studies in healthcare have demonstrated that our judgements, of staff actions and behaviours, are also influenced by what is known as 'outcome bias'.[192-195] Outcome bias involves evaluating an individual or procedure responsible for an outcome; the evaluation is considered biased when outcome information is given excessive weight.[196,197] The above studies highlight an important issue that has largely been ignored within the policy and practice of

healthcare incident investigation—that those investigating, or even consulted as part of an investigation, are potentially influenced by psychological biases—they are human too. Outside of healthcare, it is suggested that the impact of bias is broad, effecting what information is collected and how the analysis is carried out, by and with whom.[198] Bias has the potential to cause an inappropriate focus on individual culpability, and narrow or skew the exploration and understanding of an incident's causation.[199] The impact of cognitive biases is further complicated by commonly held fallacies about their nature, for instance, that they only affect corrupt, malicious or incompetent individuals, that experts are 'immune', and that simply being aware of biases allows individuals to overcome their affect.[198]

The purpose of this study was to examine the impact of outcome bias on the judgements and recommendations following hospital-based incident investigations. We also explore whether these biases might be reduced or eliminated through training or expertise in patient safety. We test the following hypotheses:

Hypothesis 1: increasing outcome severity is associated with increased judgements of responsibility and avoidability.

Hypothesis 2: increasing outcome severity is associated with increased judgements of the importance of investigation.

Hypothesis 3: increasing outcome severity is associated with more recommendations, and more punitive recommendations.

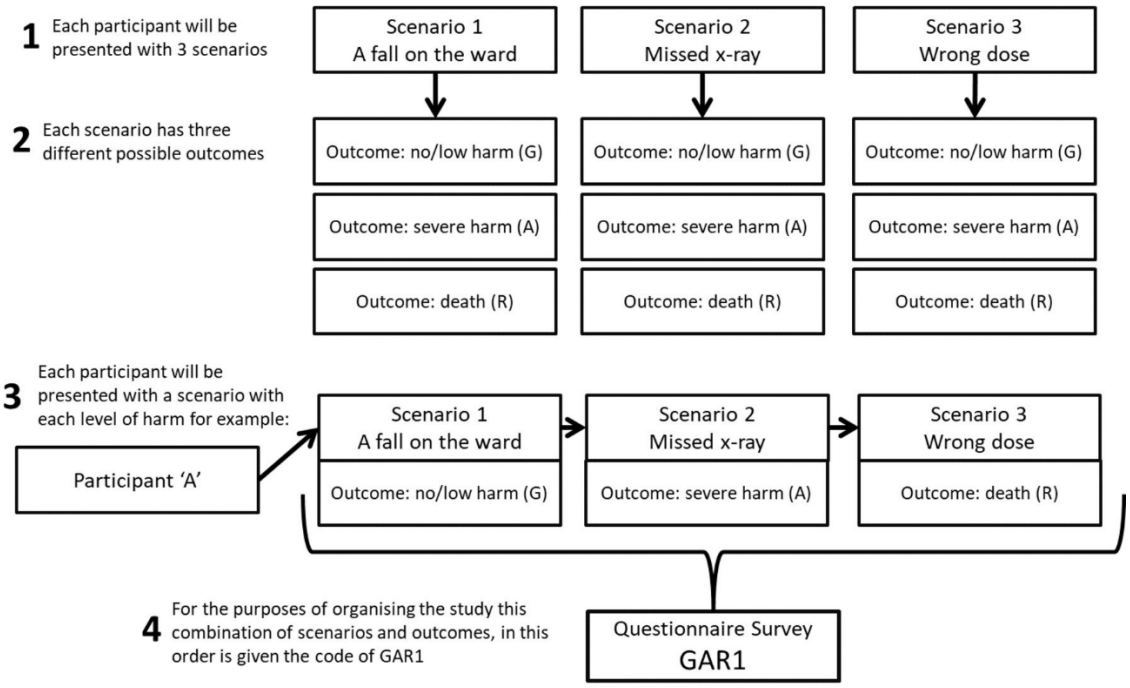
Hypothesis 4: expertise in patient safety will reduce outcome bias.

4.3 Methods

4.3.1 Study design

Using an experimental design, we developed and distributed an online questionnaire presenting three fictitious incident scenarios, along with the findings of an investigation for each (see appendix 3 for the scenarios(198) and appendix 4 for an example questionnaire(page 204)). The scenarios were based on real incidents and produced by WL, JOH, RL and CV who have a combined 86 years' experience in patient safety research, systematic review of patient safety incidents and analysis of incident investigations and recommendations. While the scenarios remained the same, regarding the events leading up to the incident and the contributing factors identified by an investigation, across three conditions, we manipulated the outcomes for the patient ((1) no/low harm, (2) severe harm, (3) death). Participants were presented with all three scenarios; one resulting in no harm, one severe harm and one that resulted in death (figure 4.1). The order in which scenarios were presented to participants was randomised, resulting in nine versions of the questionnaire, to mitigate order effects.[257,258] Repeated measures within individuals were intentionally designed to enhance the statistical power of the study, compared with a purely between-subject design. By presenting each participant with all three outcome scenarios, we control for individual differences, thus reducing variability and increasing the precision of our estimates. This approach allows us to detect smaller effects with a given sample size, as each participant serves as their own control.

Figure 4.1 Participant scenario allocation. (G = no/low harm, A = severe harm, R = death)



For each incident scenario, participants were asked to rate, on a 1–5 Likert scale, how responsible the involved healthcare professionals were in causing the incident (1=not at all, 5=entirely), how avoidable it was and how important an investigation of the incident was. Participants were also asked to select up to five (of a possible eight) recommendations to prevent incident recurrence. To verify that our manipulation of the outcomes was effective, we asked participants to identify the outcome of each scenario (no/low harm, severe harm or death) as they perceived it.

4.3.2 Participants

Participants were recruited via adverts shared on the social media platform Twitter/X. A link in the advert directed potential participants to the online questionnaire (appendix 4). The first section of the questionnaire contained questions to establish suitability for the study, which participant group they belonged to, and to gain consent. The remainder of the questionnaire presented participants with scenarios 1–3, in a random order. Following each scenario, participants were asked to answer questions as detailed

above. While participants were not given the option to save progress and complete at a later time, there was no time limit on how long participants could take to complete the questionnaire. As no name or contact information was collected from participants, there was no follow-up for completion.

Participants were recruited from three groups: (1) public, (2) healthcare staff, (3) people with expertise in patient safety or investigation of patient safety incidents.

The group of people with expertise in patient safety/investigations were further divided into those who had a clinical background (clinical experts), and those from a non-clinical background, such as researchers, policymakers or human factor engineers (non-clinical experts). Clinical expertise was defined as someone who had a background of working in healthcare (e.g., nurse, doctor, manager) and having gone on to complete at least 10 investigations, undergone patient safety or investigation training, and held a job role involving patient safety or investigation. Non-clinical expertise was defined as practical or academic job role involving patient safety or human factors in relation to incident investigation. WL and RL independently reviewed participants' answers about professional background, training and involvement in investigations, in order to assign to an expertise category. WL and RL compared allocations, any unresolved disagreements, were discussed with a third author (JOH).

4.3.3 Sample and setting

Given the innovative nature of this research, precise effect size estimates were unavailable, complicating sample size calculation.30 31 However, drawing from similar studies,[193-195] we conducted a power analysis considering plausible effect sizes (d=0.4) and variability. This analysis indicated that a minimum sample size of 150 would ensure sufficient statistical power and reliable results.

4.3.4 Analysis of recommendation choice

Following each scenario, participants were presented with eight recommendations; two punitive, followed by two weak, two medium and two strong, as defined by the action

hierarchy.[100,251,254] The recommendations were drafted by WL and reviewed and modified by JOH, RL and CV (see appendix 3 for recommendations). As well as descriptive statistics for participant recommendation choices, a weighted recommendation score was used to produce a single number representing a participants' recommendation choice. The recommendation score (RecScore) was calculated to represent degree of 'system-orientated' versus 'individual-oriented' recommendation selections. The minimum possible score was −1, representing a choice of recommendations that could be considered punitive, and the maximum possible score was 3, representing a system-focussed selection. RecScore was calculated as below:

$$RecScore = \frac{((n\ punitive\ x\ -1) + (n\ weak\ x\ 1) + (n\ of\ medium\ x2) + (n\ strong\ x3))}{Total\ number\ of\ recommendations\ selected}$$

4.3.5 Statistical analysis

In order to avoid assumptions about the equidistance of points in the Likert-Scale responses (responsibility, avoidability, importance of investigation), we produced both ordinal logistic regression models (results available on request) and multilevel linear models. Given the outcomes were similar, we opted to report the results of the linear models for simplicity.

Given the nested design (repeated measures within individuals), multivariable linear mixed (multilevel) models (MLM) were used (fit via Lmer in R), with random intercepts specified for participants.[259,260] We ran separate models for each outcome, namely, responsibility, avoidability, importance of investigating, number of recommendations (nRec) and recommendation score (RecScore).

There is evidence to suggest that age and gender, which might alter cognition and attitudes, could be important confounding factors[261-265] as well as the differences in the scenario 'story', and participant group. We also felt it reasonable to consider participants' previous involvement in incidents as a potential confounder, as a victim, staff member involved or both.

For each outcome, we built three models with an increasing number of variables to account for these potential confounding factors:

- Model 1—scenario outcome, participant age, participant gender.
- Model 2—model 1 variables+scenario, participant group.
- Model 3—model 2 variables+participant previous incident involvement (none/victim/staff member).

As well as coefficient estimates, several statistics were produced from the models to evaluate model fit. These included (1) the intraclass correlation coefficient (ICC), providing an estimate of the proportion of total variance in each outcome that was attributed to the grouping structure in the data (ie, between participant differences), (2) R2Marginal, representing the proportion of variance explained solely by the fixed effects, disregarding the random effects: gauging how well predictors elucidate the outcome variable, excluding the multilevel structure's consideration and (3) R2Conditional, representing the same as R2Marginal but also including the variance explained by the random effects.

We also produced Bayesian Information Criterion (BIC) statistics, which allowed us to compare model fit between models specifying the same outcome (lower scores=better fit). Primarily these were used to gauge whether adding further predictors in model 2 or 3 improved model fit. Subgroup analysis was performed to calculate mean differences in responsibility ratings, nRec and RecScore between the participant groups. The correlation between designed scenario outcome and participant reported outcome was calculated to check the manipulation. Throughout the manuscript, we specified alpha at 5% (two tailed).

4.4 Results

Two hundred and twelve participants completed the questionnaire (table 4.1), resulting in 636 observations (three observations per participant). Missing data were low, 41 of 5936 data items (0.69%). Participants were mostly women (n=166, 78.3%) and had an average age of 44 (range 17–80, SD=13.6). Members of the public made up the largest group (n=100, 47.2%), followed by healthcare staff (n=71, 33.5%) and experts (n=41, 19.3%; clinical 30; researcher 11). Approximately a third of participants had been a victim of a safety incident (either personally or a close relative) (n=63, 30.0%), or involved as a member of staff (n=70, 33.0%). A small proportion of participants had been both a victim and a member of staff involved in a safety incident (n=19, 9.0%).

Table 4.1 Participant characteristics

	n	%
Public	100	47.2%
Healthcare Staff	71	33.5%
Experts	41	19.3%
Female	166	78.3%
Male	45	21.2%
Other/Non-binary	1	0.5%
Previous Incident Involvement		
None	98	46.2%
Victim	44	20.8%
Healthcare Staff	51	24.1%
Both	19	9.0%
total	**212**	

4.4.1 Manipulation check

Designed outcome severity was highly correlated with participant-reported severity (scenario 1: $r=0.967$, $p<0.001$, scenario 2: $r=0.912$, $p<0.001$, scenario 3: $r=0.936$, $p<0.001$).

4.4.2 Multilevel modelling

When reporting results of MLM below, we refer to those obtained from the model with the lowest BIC, for each outcome. The difference between BIC values across the models was not significantly different and R2c values ranged from 40% to 54%. The ICC ranged from 26.8% to 50.9% across models indicating a significant degree of clustering, supporting the use of MLM.

See appendix 5, page 210, for full model results. Across the models, there appeared to be no significant effects for the adjustment variables age or gender. There were significant effects for scenario (fall, X-ray, wrong dose) in all outcome variables; in other words, the details of the incident scenarios (eg, events and people involved) had effects on participants' judgements and responses to all questions. Having been a victim of an incident (personally or close family member) had a significant effect on responsibility ratings ((β=0.50 (CI 0.20 to 0.83), p=<0.001) and importance of investigating ((β=0.31 (CI 0.04 to 0.57), p=<0.001). Having been a staff member involved in an incident before appeared to have no significant effects on the outcome variables.

4.4.3 Hypothesis 1: increasing outcome severity is associated with increased judgements of responsibility and avoidability

This hypothesis was partly supported. As the outcome for the patient in the scenario became more severe, participants judged the staff involved in the incident as more responsible for causing it. This is demonstrated by the increasing responsibility rating means for no/low harm (2.99), severe harm (3.16) and death (3.25) in table 4.2. These means are adjusted for age, gender, scenario and participant group. Multilevel modelling demonstrated a significant association between outcome severity and responsibility ratings, the most significant difference noted between death and no/low harm ((β=0.26 (CI 0.11, 0.41), p≤0.001) (table 4.2 and appendix 5).

Table 4.2 Mean response ratings. by outcome and participant group, with 95% CIs

Table 2 **M*ean Response Ratings*** By Outcome And Participant Group, With 95% Confidence Intervals							
	Outcome			Participant Group			
	no/low-harm	Severe harm	Death	Public	Staff	Clinical Experts	Non-clinical experts
Responsibility	2.99 (2.78,3.20)	3.16 (2.95,3.37)	3.25 (3.04,3.46)	3.67 (3.46,3.88)	3.39 (3.17,3.61)	2.82 (2.48,3.16)	2.66 (2.11,3.2)
Avoidability	4.02 (3.84,4.2)	3.98 (3.8,4.16)	3.98 (3.8,4.16)	4.31 (4.14,4.48)	3.94 (3.76,4.12)	3.78 (3.5,4.06)	3.93 (3.48,4.38)
Importance of investigating	4.09 (3.92,4.27)	4.48 (4.3,4.65)	4.73 (4.56,4.9)	4.43 (4.26,4.6)	4.54 (4.36,4.72)	4.39 (4.11,4.67)	4.37 (3.93,4.82)
Number of Recommendations	3.63 (3.4,3.86)	3.56 (3.34,3.79)	3.83 (3.6,4.06)	3.82 (3.59,4.04)	3.92 (3.68,4.16)	3.44 (3.07,3.81)	3.53 (2.94,4.11)
Rec score	2.04 (1.94,2.13)	1.97 (1.87,2.07)	1.96 (1.86,2.06)	1.84 (1.71,1.96)	1.81 (1.71,1.91)	2.07 (1.91,2.24)	2.23 (1.99,2.47)

Participants were asked to rate avoidability of the incident. Findings in table 4.2 and the multilevel model show little difference in the ratings of avoidability for the different outcomes of low harm (4.02), severe harm (3.98) and death (3.98). All ratings were high, suggesting that irrespective of outcome, participants considered these incidents to be highly avoidable.

4.4.4 Hypothesis 2: increasing outcome severity is associated with increased judgements of importance to investigate

This hypothesis was supported. All ratings were above 4 on the five-point scale. However, when the outcome for the patient was death, the mean score for importance of investigation was 4.73 compared with 4.48 for severe harm and 4.09 for no/low harm. Multilevel modelling (appendix 5) confirmed a statistically significant association between outcome severity and importance of investigating, with the biggest difference observed between no/low-harm and death (β=0.63 (CI 0.50, 0.76), p=<0.001) (table 4.2 and appendix 5).

4.4.5 Hypothesis 3: increasing outcome severity is associated with selecting more recommendations and more punitive recommendations

Participants were asked to select up to 5 recommendations per incident and 2452 recommendations were selected across 636 incidents. There was no significant observed differences between the nRec selected for no/low harm incidents (average 3.81, SD 1.05), severe harm (3.74, SD 1.21) or death (4.01, SD 1.05). While the models did demonstrate a statistically significant increase in nRec selected for death outcome versus no/low harm outcome, the difference was very small (β=0.20 (CI 0.02, 0.38), p=0.03), representing a fifth of a recommendation (table 4.3 and appendix 5)

Table 4.3 illustrates the types of recommendations selected, categorised as either punitive (n=154, 6.3%), weak (n=682, 27.8%), medium (n=910, 37.1%) or strong (n=706, 28.8%), in terms of their likelihood of improving safety (see online supplemental appendix 1 for recommendations).[100,254] It is important to highlight that punitive recommendations made up 8% of those selected when the outcome for the patient was

death, 6% for severe harm and 5% for no/lo harm, suggesting that punitive recommendations are more likely to be selected when the outcome for the patient is worse.

Table 4.3 Number and types of recommendations selected

	Number of recommendations selected within each category					average total recommendations selected by each participant
Patient outcome severity	Punitive (1+2)	Weak* (3+4)	Medium* (5+6)	Strong* (7+8)	Total	
Death	67	227	311	246	851	4.01 (sd 1.05)
Severe harm	45	231	295	222	793	3.74 (sd 1.21)
No/Low harm	42	224	304	238	808	3.81 (sd 1.14)
total	154	682	910	706	2452	
	6.3%	27.8%	37.1%	28.8%		
* as defined by the action hierarchy (NPSF 2021).						

Multilevel modelling demonstrated that mean recommendation scores, across participants groups, reduced as the patient outcome became more severe, indicating a more individual-focus to recommendation choices. RecScore was lower when the outcome for the patient was death versus no/low harm and severe harm versus no/low harm, but the differences were not statistically significant ((β=−0.08 (CI −0.16, 0.00, p=0.057) and (β=−0.06 (CI −0.15, 0.02, p=0.137)).

4.4.6 Hypothesis 4: expertise in patient safety will reduce outcome bias

Our results suggest that those with non-clinical or clinical expertise in safety assign less responsibility to staff for causing an incident than staff (difference=0.73 (95% CI 0.14 to 1.32) and 0.57 (95% CI 0.18 to 0.96)) and the public (difference=1.01 (95% CI 0.44 to 1.58) and 0.85 (95% CI 0.48 to 1.22)); with no significant difference between the public and staff (difference=0.28 (95% CI −0.01 to 0.57))(table 4.4). Furthermore, those experts from a non-clinical background appear to assign less responsibility to staff than those from a clinical background (tables 4.2 and 4.4, table 4.4).

Table 4.4 Differences between participant groups

Comparison	Responsibility				Avoidability				Number of Recommendations				Recommendation Score			
	Difference	**95% CI-**	**95% CI+**	**p**	**Difference**	**95% CI-**	**95% CI+**	**p**	**Difference**	**95% CI-**	**95% CI+**	**p**	**Difference**	**95% CI-**	**95% CI+**	**p**
Public vs Staff	0.28	-0.01	0.57	0.056	**0.37**	0.13	0.61	0.002	-0.10	-0.41	0.21	0.51	-0.05	-0.15	0.05	0.338
Public vs Clinical Experts	**0.85**	0.48	1.22	<0.001	**0.53**	0.22	0.84	< .001	0.38	-0.01	0.77	0.06	**-0.32**	-0.46	-0.18	< .001
Public vs Non-Clinical Experts	**1.01**	0.44	1.58	<0.001	0.38	-0.09	0.85	0.108	0.29	-0.32	0.90	0.35	**-0.35**	-0.55	-0.15	< .001
Staff vs Clinical Experts	**0.57**	0.18	0.96	0.006	0.16	-0.17	0.49	0.335	**0.48**	0.05	0.91	0.03	**-0.27**	-0.41	-0.13	< .001
Staff vs Non-Clinical Experts	**0.73**	0.14	1.32	0.015	0.01	-0.46	0.48	0.973	0.39	-0.24	1.02	0.22	**-0.30**	-0.50	-0.10	0.004
Clinical Experts vs Non-Clinical Experts	0.16	-0.47	0.79	0.614	-0.15	-0.66	0.36	0.560	-0.09	-0.76	0.58	0.80	-0.03	-0.25	0.19	0.763
Means that were produced to calculate between mean differences were adjusted for age, gender, scenario, and outcome.																

Table 4.5 Number of participants selecting punitive recommendations by participant group

	Death		Severe harm		No/Low harm	
Group	n	% of group	n	% of group	n	% of group
Public	34	34.0%	27	27.0%	25	25.0%
Healthcare staff	18	25.4%	9	12.7%	9	12.7%
Experts	10	24.4%	5	12.2%	6	14.6%

We observed no difference in mean avoidability ratings between staff and experts (table 4.2), but we observed a small difference between both the public and staff (0.37 (95% CI 0.13 to 0.61)) and public and clinical experts (0.53 (95% CI 0.22 to 0.84))(table 4.4). Our results suggest that those with in-depth knowledge of the clinical environment (staff and clinical experts) may perceive incidents as less avoidable than the public or non-clinical experts.

We observed no difference between participant groups in ratings of importance of investigating the incidents within the vignettes, with mean ratings ranging from 4.37 for non-clinical experts to 4.54 for healthcare staff (table 4.2).

There were no significant differences in the total nRec that participants from different groups selected (table 4.2 and table 4.4). We did, however, observe differences in the types of recommendations that were selected. Those with patient safety expertise are likely to select stronger recommendations, according to the AH[100,254,258] than members of the public or healthcare staff, with the biggest differences in recommendation score (scored between −1 and 3) between non-clinical experts and the public (0.35 (95% CI 0.15 to 0.55)) and clinical experts and the public (0.32 (95% CI 0.18 to 0.46))(table 4.4). It is important to highlight the differences observed in the selection of punitive recommendations (table 4.5). The percentage of public selecting punitive recommendations increased from 24%, at the no/low harm level, to 27% and 34% at severe and death harm levels, respectively. On the other hand, 12.2%–14.6% of staff and experts selected punitive recommendations for no/low harm and severe harm, which increased to 25.4% and 24.4% when the outcome for the patient was death. In other words, more staff and experts selected punitive recommendations when the patient outcome was worse.

4.5 Discussion

The results of this study show that outcome knowledge is associated with changes in how individuals judge and respond to incidents, irrespective of their background (public/professional/expert) or previous experiences with incidents (harmed or involved). While this association has been demonstrated in other domains, to the

authors' knowledge, this is the first study to examine this issue specifically in the context of healthcare incident investigations.

Some of the absolute effects on responses, in our results, appear small. Despite this, we propose even small effects on responses could have a significant impact on patient safety investigations. First, we have demonstrated the effect of outcome bias on three aspects of the investigation process (whether to investigate, the responsibility of staff and the selection of recommendations). Rather than considering the impact of each individually, we need to consider the cumulative impact of outcome bias that may occur at repeated time points, on multiple decisions and the many people involved through the course of a single investigation.[198] Second, given 2–3 million incidents are reported in England alone even a very small impact on each incident investigation has far-reaching consequences at a national level.

4.5.1 Increasing outcome severity is associated with increased judgements of responsibility but not avoidability

Our results suggest that the more severe the outcome of an incident, the greater the responsibility that will be assigned to the staff involved. This effect of outcome-knowledge bias on responsibility has been demonstrated many times in different contexts,[190,191] and so it is not surprising to find it within the context of healthcare incident investigations. Thus, while Dekker and others highlight the need for a shift in responsibility for patient safety from human error to symptom of trouble within a system,[266] unrecognised outcome bias may be hindering progress in this direction.

While avoidability did not appear to be associated with worsening outcome severity, our results suggest participants considered all the incidents relatively avoidable, with mean ratings across all outcome severities and group categories being approximately 4 out of 5. These generally high scores may be the result of hindsight bias where once the outcome is known people tend to alter their perception of how likely an event was to occur, sometimes referred to as the 'knew-it-all-along' effect.[267] While both hindsight bias and outcome bias involve the projection of new information into the evaluation of past events or actions, hindsight bias involves the denial that outcome information has

influenced judgements.[268-270] As we did not ask participants how they came to their judgements, we have not specifically examined hindsight bias.

Hindsight and outcome biases may cause investigators to focus on poor decisions or missed opportunities rather than other factors, in incident causation.[267,271,272] The avoidability or preventability of an incident may be described as complex[273] and beyond the scope of this paper.

4.5.2 Increasing outcome severity is associated with increased judgements of importance to investigate

Our results suggest people consider it more important to investigate an incident when a patient comes to greater harm. There are a number of purported reasons for carrying out investigations: to improve the safety and quality of care; to assign accountability; to litigation or compensation, for restoration, or to repair or protect organisational reputation.

This study was designed, so that the events leading up to the incident and the systems in which the incidents occurred were identical; it was by chance that each of the alternative outcomes occurred (no/low harm, severed harm and death). Therefore, it could be argued that the 'opportunities for learning' were the same, no matter the outcome for the patient. When allocating resources for investigation, organisations are encouraging more proportionate responses and a move away from responses based on subjective thresholds and definitions of harm.[274] Our study demonstrates that level of harm remains an important factor in how people decide how to investigate incidents. New policies and frameworks alone might struggle to overcome this especially in jurisdictions that mandate investigations for higher harm incidents, thus legitimising outcome bias in health policy.[70,275]

4.5.3 Increasing outcome severity is not associated with more recommendations but is associated with more punitive recommendations

While there was no observed difference in the number of recommendations selected between levels of harm or participant groups, there was an overwhelming tendency to select recommendations than not (an average of 4 out of 5 were selected for all three

scenarios). Adams *et al* demonstrated that humans prefer additive change than subtractive, for example, adding a checklist rather than removing one.[276] While we cannot be sure this is the case in our study, our results imply that safety investigations could contribute to the creation of safety clutter or low-value safety practices.[277,278]

Our results suggest that knowledge of a severe patient outcome, alone, may increase the chance of punitive recommendations being selected. This inappropriate focus on individuals may distract improvement efforts away from a system approach, such as redesigning processes and equipment. The impact of outcome bias also has the potential to contribute to the already present culture of blame within healthcare, hampering efforts to improve patient safety and encouraging the adoption of potentially harmful defensive practice among clinicians.[279,280]

4.5.4 Expertise in patient safety can reduce some biases

Our results demonstrate differences in how people from different backgrounds and expertise respond to incidents. Expertise appears to mitigate perceptions of responsibility and the reduce the selection of punitive recommendations but not fully. When the patient outcome was death, the proportion of experts and staff selecting punitive recommendations doubled. This suggests that clinical knowledge or patient safety expertise alone, may be useful but insufficient to mitigate against outcome bias. This is perhaps expected given research showing that a number of factors affect the presence or impact of biases, such as expertise, previous experience, cognitive ability, bias awareness, tolerance of ambiguity and organisational culture.[272,281,282] The interesting differences observed between clinical and non-clinical experts suggest that the type and origin of expertise are important in ensuring effective investigations. Further research should focus on understanding the impact of investigator cognition and bias as well as other factors such as professional background and training on investigations.

4.6 Implications for policy and practice

We highlight the different and sometimes conflicting responses of the public, staff and experts, which will need to be considered within policy and future research. While patient harm plays an important part in individuals' responses to an incident,

policymakers will need to consider when harm severity is justification for investigation and when it is not. Those with responsibility for oversight of investigations and investigators should have an understanding of bias and how it might impact investigations. Future research should continue to develop and empirically test strategies to mitigate the impact of cognitive biases on investigations, such as investigator training, blinding and unmasking (eg, to patient outcome) and independent verification.[272] Further research is needed to understand not only the impact of bias on investigation but also the increasing number of alternatives to traditional investigation in patient safety, such as After-Action Review, 'SWARM' and Structured Judgement Review.[82,283] Policymakers and organisations should ensure that investigations are led by those with expertise and experience and would benefit from defined competencies for investigators.

4.7 Limitations

Our study has a number of limitations. We examined outcome bias in individuals; but investigations may be carried out by a team; future research is needed to understand how investigations performed by a group may or may not mitigate the impacts of outcome bias, or indeed introduce other cognitive and social issues, such as 'Group Think' or shared information bias, on the entire life cycle of an investigation.[284-286] While a team of investigators is recommended, in our experience and in discussion with a wide range of stakeholders, we suggest that not all investigations are carried out by teams. The most 'serious' incidents might well be, but many other less serious will be carried out by an individual. Even when a team is reported to have carried out an investigation, it is likely that a single lead will have carried out most of the analysis. Though the scenarios were based on real incidents, they might not fully capture the complexity of actual events, limiting the generalisability. Participant recruitment was via X, (formerly Twitter), which may not represent the broader population. We did not account for other participant characteristics such as professional backgrounds of staff participants (eg, nursing, medical, etc), workplace (urban large academic centre vs rural settings, etc), health policies relevant to incident reviews in their jurisdiction or cultural backgrounds, which might influence how individuals respond to the question in this study.[194,287] This study was not powered to detect differences in other participant

grouping. Our MLMs suggested we identified a significant number of confounding factors; there are likely to be several other confounding or moderating factors that we have not explored, which impact the decisions and judgements of people investigating incidents. Identifying and exploring the impact of these factors might be of interest for future research.

4.8 Conclusion

This study adds to the body of evidence that human cognition and bias are likely to have a significant impact on how incidents are investigated within healthcare. We highlight the conflicting views of the public, staff and experts; the difficult topic of individual responsibility, accountability and 'blame' and ultimately the appropriate distribution of 'causality' of a systems performance between individuals and systems.

Chapter 5: (Study 3) What does good look like in healthcare incident investigations and recommendations? A modified Delphi study.

Drafted for submission but not yet submitted. William Lea (WL) developed the study design with support from JOH, RL and CV as well as an external expert steering group (Claire Cox(CC), Sarah Seddon(SS) and Paul Bowie(PB)). WL led every step of the research with support JOH, RL, CV, CC, SS and PB. WL drafted the manuscript, and edited with input and guidance from JOH, RL, and CV.

5.1 Introduction

It is estimated that one in ten patients will experience an 'adverse event' (AE), that is something going wrong in the way in which care is delivered.[13] These events are often investigated and recommendations made, with the intention of preventing recurrence or improving safety. There are, however, increasing concerns that these recommendations and preceding investigations are not contributing to improved safety, and potentially contributing to safety clutter.[9-12,277,278,254] In a recent review, the authors highlighted a surprising lack of data pertaining to the implementation of recommendations, evaluation of their effectiveness or even recording the recurrence of incidents that originally triggered investigations.[254] There is no clear agreement on how recommendations should be judged for 'effectiveness' and despite its widespread use to presumptively evaluate recommendation effectiveness, the Action Hierarchy (AH) lacks empirical evaluation and adaptation within healthcare.[213,234,236,237,254]

In their 2017 article on the shift from 'what hospitals learn' from incidents to 'how hospitals learn', Leistikow and colleagues argued that evaluating the investigation process may provide an alternative to measuring outcomes from recommendations.[211] The 'quality' of the investigation process itself and the way in which recommendations are generated may provide a useful way to evaluate the way in which investigations contribute to improved safety.[15,107,211]

Given the identified links between the investigation and the recommendation generation process, a reliable way of assessing quality for both would be needed. There

have been several attempts to identify criteria to evaluate the quality of investigations or investigation approaches[107,211,288,289] but there remains a lack of empirical basis for this and consensus about judging the quality of investigations or recommendations in healthcare.

The aim of this study was to seek consensus on what 'good' looks like in investigation and recommendation generation. More specifically, we aimed to identify criteria that would indicate a good investigation and recommendations. As briefly mentioned previously the recommendations generated from investigations are likely to be linked to the findings and therefore quality of the investigation as a whole. Therefore some of the criteria, within this Delphi, may relate to the investigation process leading up to recommendation generation based on the assumption that a poor investigation may result in poor recommendations.

The purpose of developing criteria is to inform evidenced-based evaluation of the process of recommendation generation, including the preceding investigation approach. This could be in the context of practice, by hospital patient safety teams or investigators, as well as researchers and those responsible for policy.

5.2 Methods

5.2.1 Study Design

Guided by an expert steering group, we conducted a 2-stage consensus-building approach: 1) identifying and drafting candidate criteria that would indicate a good investigation or recommendation; 2) conducting a 3-round modified Delphi process to identify, refine wording and to gain consensus on candidate criteria.[290-294]

5.2.2 Steering group

The experience of the first author and steering group (JOH, RL, CV, SS, CC and PB) were also used in drafting candidate criteria. The authors had expertise in patient safety research as well as systematic review of patient safety incidents and analysis of incident investigations and recommendations, both in clinical and academic contexts. The steering group brought experience at the local organisational (e.g. hospital) as well

as regional or national level. Importantly, one member of the steering group had personal involvement as a parent in an incident and subsequent investigations. The possible criteria were drafted by WL and refined following review by the steering group. To account for possibility that criteria might have been missed through this process, Delphi panel participants would be asked for other suggested criteria in round 1.

5.2.3 Stage 1: Developing draft quality criteria

A list of possible draft criteria was generated that could be entered into the modified Delphi study.

5.2.3.1 Sources

Firstly, the existing academic and grey[295] literature on the practice of investigation and recommendation generation was examined. Initially existing criteria sets,[296-300] and policies or publications providing guidance on how to undertake investigations or make recommendations were identified.[98,107,124,183,251,301] Existing criteria sets provided an important starting point for drafting the criteria in this Delphi.

Articles referenced by these initial sources were then identified as well as other sources recommended by members of the steering group. Finally the reference lists of identified articles or sources were examined to identify other relevant articles. Articles or sources were read to identify features of investigations or recommendations that were mentioned either as good or bad practice, within the documents. Articles cited by the initial list were added, until this ceased to yield further candidate criteria. The final list of documents reviewed(46) comprised 34 journal articles,[11,12,144, 145,159,175,179,180,218,236,266,272,302-304,306-323] six white papers/reports,[107,160,324-327] five policies/guidelines,[103,183,300,301,328] and one PhD.[329]

5.2.3.2 Forming the draft criteria

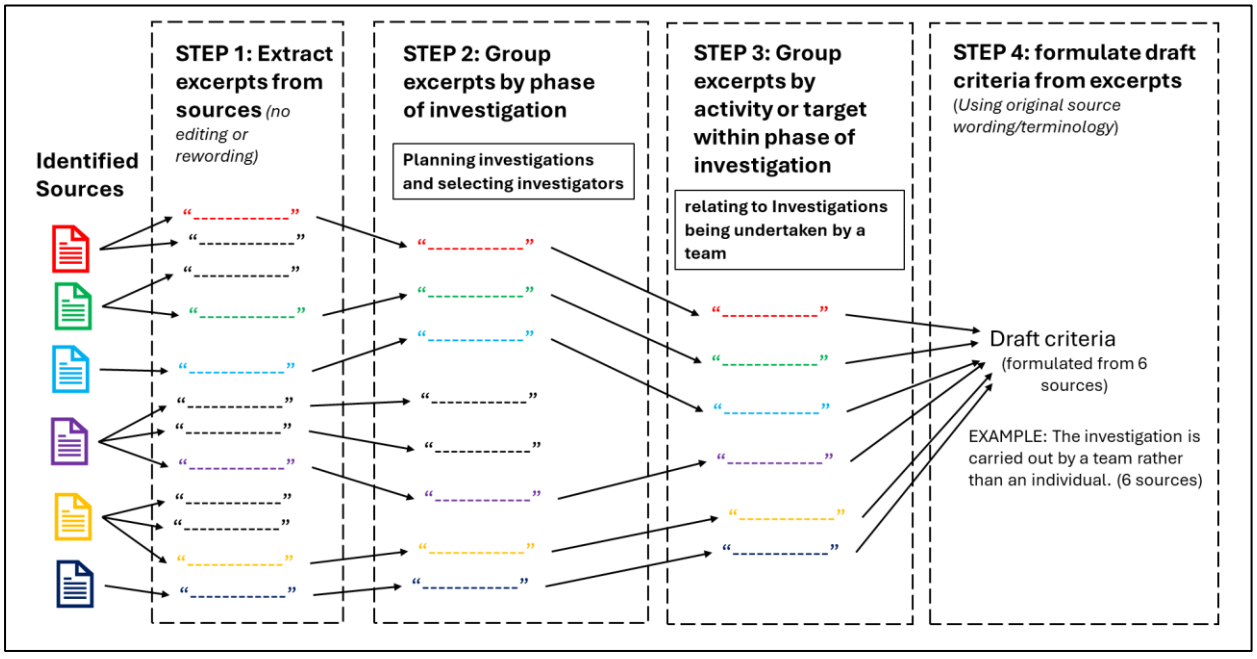
Draft criteria were formulated following the steps illustrated in figure 5.1 below. The grouping of excerpts and formulation of draft criteria was provisionally carried out by WL and then reviewed by JOH, RL, CV, SS, PB and CC.

Step 1: 496 excerpts of text, relating to the quality of investigations or recommendations, were extracted from the sources above. The text was extracted without editing, re-wording or summarising.

Step 2: The text excerpts were organised into one of six phases of the investigation process:[107,183]

1) Planning investigations and selecting investigators (86 excerpts from 22 sources),
2) Engaging people in the investigation (64 excerpts from 19 sources),
3) Using a systems approach to gathering data, information and evidence (168 excerpts from 39 sources),
4) Developing recommendations (120 excerpts from 29 sources),
5) Documenting the investigation (30 excerpts from 12 sources), and
6) Follow-up of the investigation and recommendations (28 excerpts from 9 sources).

Figure 5.1 Process for forming draft criteria



Step 3: Excerpts grouped into investigation phases were then grouped into the activity or characteristic which they described, for example the excerpts below all related to the investigation being carried out by a team:

"RCA2 teams should be composed of 4 to 6 people."[183]

"Investigations should be undertaken by a small nominated team"[314]

"The investigation team should consist of three or four people facilitated by the investigation leader. It is important to identify team members with multiple skills and the time to commit to the process. For very serious incidents the team may need leave from 'normal duties' to focus on incident investigation and analysis."[107]

"Learning responses are not undertaken by staff working in isolation. A learning response team should be established to support learning responses wherever possible."[300]

"Investigations must be led by trained investigators with the support of an appropriately resourced investigation team...."[327]

"Does your organisation use a team approach to investigation (versus a single investigator)? Where a single investigator model is used, your organisation increases the risk that investigations will be based largely on the assumptions and interpretation of one person."[324]

Step 4: Draft criteria were formulated from the organised excerpts. Careful attention was made to retaining the wording and terminology used in the original sources.

A total of 92 criteria were formulated; 21 relating to planning investigations and selecting investigators, 13 to engaging people in the investigation, 24 to using a systems approach to gathering data, information and evidence, 17 to developing recommendations, 12 to documenting the investigation, and 5 to follow-up of the investigation and recommendations.

5.2.4 Stage 2: Modified Delphi Process

A modified Delphi technique,[290-294] was employed to gain consensus, amongst an expert panel, on how well criteria would indicate a good investigation or recommendation. Consensus on the wording of criteria would also be sought, to ensure they 'made sense'. The Delphi was conducted over 3 rounds and a 5-month time period.

5.2.4.1 Expert Panel Participants

Participants were invited to the expert panel if they met one or more of the following inclusion criteria:

- Public experience of investigations or those directly affected by them (having been the victim or close relative/person with caring role for the victim of a patient safety incident and subsequently involved in the investigation)[223]
- Practical experience of carrying out investigations (defined as having completed >10 investigations)
- Human factors and safety expertise (defined as having held an academic or practical job role in these areas for >12 months)
- Experience of designing or implementing interventions and system change, following investigations (defined as having held an academic or practical job role in these areas for >12 months, which may include quality improvement)*
- Responsibility for investigation processes or policy within local organisations or nationally (defined as having held a relevant job role for >12 months)

(*Examples might include patient safety leads, quality improvement leads, senior clinicians in hospitals)

The aim was to recruit 80 participants in order to ensure at least 40 participants were retained throughout the three Delphi rounds, similar to other studies.[291,294,330] Within this, the aim was to recruit a minimum of 5 participants with expertise from each of the above categories. It was anticipated that some participants may have expertise from more than one category.

5.2.4.2 Expert panel recruitment

Potential panel experts were identified in two ways. Firstly, as personal invitation has been shown to improve recruitment and retention,[331] experts were suggested by the steering group, and were invited to take part via email. The email contained a participant information sheet and link to an online form to collect information about expertise and gain consent. Initial contacts were asked to identify and forward the invitation email to others who might be suitable/willing to take part (snowball sampling). Secondly, experts

were identified through academic publications and by sharing a recruitment email through gatekeeper contacts at organisations such as the Health Services Safety Investigations Body and Harmed Patient Alliance.

5.2.4.3 Delphi Round 1

A structured questionnaire, hosted on app.onlinesurveys.jisc.ac.uk, was distributed to the experts. Experts were given one month to complete each round and were able to save progress if they did not want to complete the questionnaire in one sitting. Experts were presented with each criteria, drafted in stage 1, and asked to rate how strongly they disagreed or agreed with the statement "This would indicate a good investigation." The number of sources from which each criteria was drafted in stage 1, was included, but not the sources themselves. Experts where then asked to respond 'Yes' or 'No' to "This criteria is written well". It makes sense." If experts selected "No" they were provided with a free text box to suggest how it could be improved. At the end of the questionnaire experts were asked whether they wished to suggest any other criteria and provided with a free-text box to provide details, and also a free-text box for any other comments relating to Round 1.

As in previous studies a nine-point scale was used with anchors of "strongly disagree" (rating = 1), "uncertain" (rating = 5) and "very strongly agree (rating = 9).(van der Scheer 2021; Taylor 2016) Across expert responses, the median was used to indicate the strength of agreement (7-9 as strong, 4-6 as moderate and 1-3 as weak), and the mean absolute deviation from the median (MADM) was the level of agreement (>1.41 as low, 1.08-1.41 as medium, and <1.08 as high).[332] The median represents an 'average' rating and the MADM gives an idea of how 'spread out' peoples' responses are; another way to describe this is as the median indicating how strongly the participants feel about the statement, and the MADM indicating how much the participants agree on this feeling.

Consensus on the wording was set at >75%; if >75% of participants select 'no' the wording would not be changed, if 75% of participants select 'yes', the steering group would review free text comments and draft a re-worded criteria for the next round.

Any criteria with a median of 7-9, and MADM of <1.08, and >75% wording agreement were not fed back into Round 2. Any criteria not meeting these thresholds were

reviewed, along with panel comments/suggestions, by the steering group and a decision made as to whether the criterion were 1) fed into Round 2 unchanged, 2) re-worded and fed into Round 2, 3) removed from the process. The steering group also reviewed the expert-suggested additional criteria to establish if unique criteria could be added to Round 2. Experts were sent a summary of the Round 1 results, prior to Round 2, so that they could see how their responses compared to that of the group.

5.2.4.4 Delphi Rounds 2 and 3

Rounds 2 and 3 followed the same structure as Round 1 but only presented those criteria that had not achieved consensus in the previous round.

5.3 Results

81 experts were recruited, 62 of whom completed Round 1, with 49 and 46 completing Rounds 2 and 3 respectively. Table 5.1 shows participant numbers and expertise at each round.

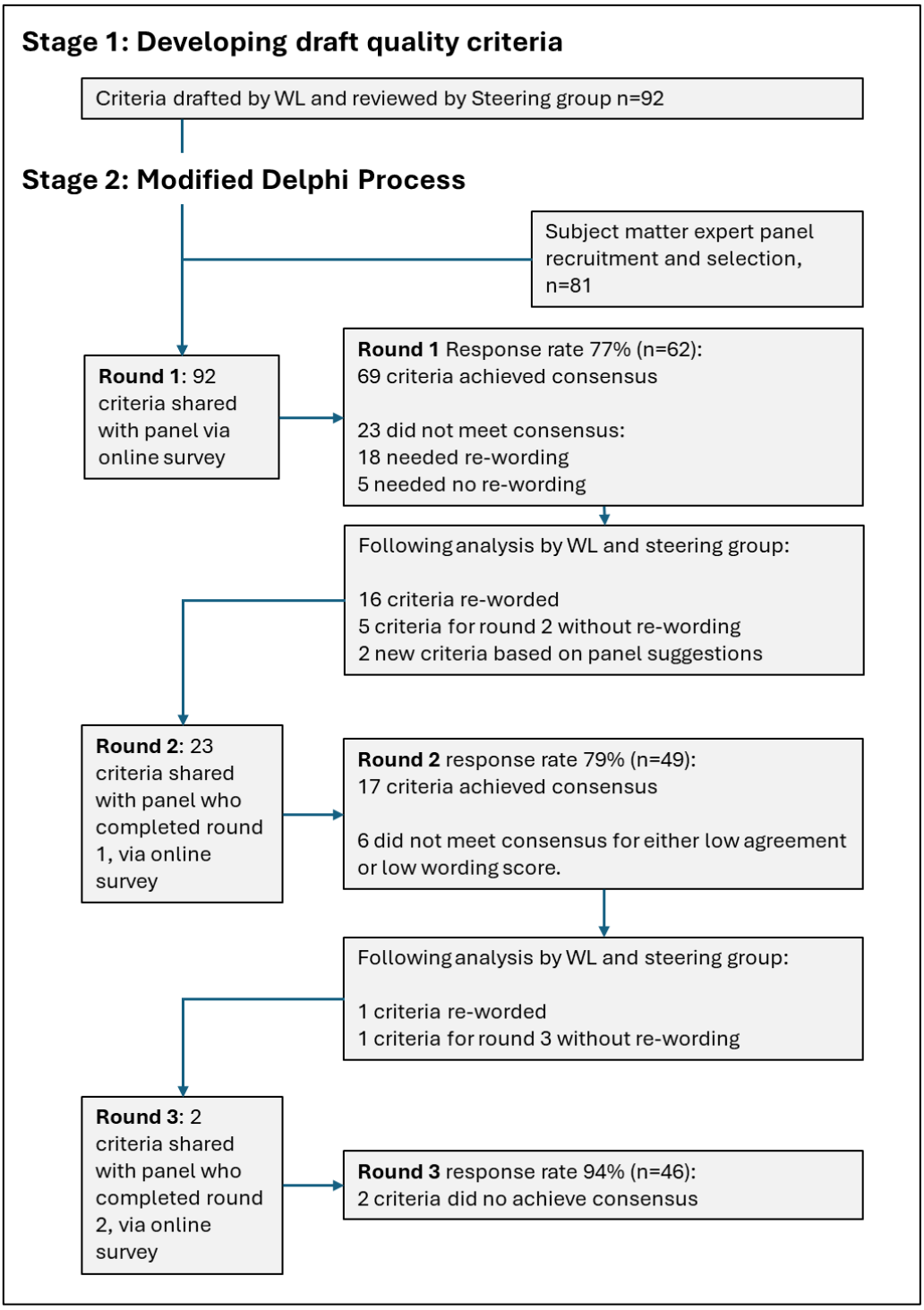
Table 5.1 The Expert panel (Expertise and Delphi Round Retention)

	Recruited		Completed					
			Round 1		Round 2		Round 3	
total experts	81		62		49		46	
Number(n) and percentage(%) of the expert panel with expertise								
Category of Expertise	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Public experience of investigations or first victims	15	19%	12	19%	9	18%	9	20%
Practical experience of carrying out investigations	47	58%	38	61%	33	67%	31	67%
Human factors and safety expertise	73	90%	57	92%	45	92%	42	91%
Experience of designing or implementing interventions and system change, following investigations	42	52%	32	52%	25	51%	24	52%
Responsibility for investigation processes or policy within local organisations or nationally	59	73%	46	74%	39	80%	36	78%
Note: many panellists had expertise from more than one category, and therefore the sum of the values under the columns labelled 'n' equal the number of panellists with the specified expertise, not the total number of panellists completing each round.								

The results are summarised in figure 5.2 below. During stage 1 92 draft criteria were identified that would indicate a good investigation or recommendation (Tables 5.2 and 5.3). At the end of Round 1, 69 criteria achieved consensus and were not entered into Round 2. Of the remaining criteria 21 criteria entered into round 2 (the wording was changed in 16 criteria, and wording unchanged in 2), 2 were removed(Criteria 72 and 81). Two new criteria were drafted, based on expert suggestions, and fed into Round 2, resulting a total of 23 for review in Round 2. At the end of Round 2 17 criteria achieved

consensus. Of the remaining criteria, 2 criteria were fed into Round 3 (1 re-worded, 1 unchanged), and four were not(Criteria 20, 21, 42, 66). At the end of Round 3 the remaining criteria had not achieved consensus (Criteria 32 and 56). Table 5.2 displays the ranked criteria, table 5.3 those which were removed with rationale.

Figure 5.2 Delphi Flow Chart



5.3.1 Use of panel comments between rounds

The steering group reviewed criteria that did not meet consensus and based on panel comments decided whether to re-word or remove from the process. We will not present all panel comments but provide below a few examples to illustrate the rationale:

Example of a criterion that was re-worded and entered into subsequent Delphi round

> *Criterion 2: The investigation team is made up of people from different professional backgrounds. (Wording used in round 1)*

This criterion achieved a median of 8 and MADM of 1, in the first round, but a wording score of 72.1%. A number of panellists suggested defining 'which backgrounds' as well as the potential for the use of 'professional' to exclude important stakeholders:

> *""Professional" limits to recognised professions. Perhaps just backgrounds without a preceding word."*
>
> *"I wonder if you need to define the professional backgrounds. I think HCAs and Porters can often provide a rich source of intelligence but are not considered professionals."*
>
> *"I think it should say it is made up of people who are capable of providing different perspectives, as this may include non- professional staff as well as professionals, depending on the type of investigation."*

Based on these and other comments, this criterion was re-worded, as below, and achieved consensus in round 2, with a wording score of 79.6%.

> *Criteria 2: The investigation team is made up of people from different professional and non-professional backgrounds to bring a range of perspectives and experiences relevant to the incident.*
> *Further explanation: By 'different backgrounds' we mean nurses, doctors, HCAs, porters, administrative staff, volunteers etc. People who are capable of providing different perspectives, which may include non-professional staff as well as*

professionals, depending on the type of investigation. This is to increase the chance of diverse perspectives being captured and a range of skills and expertise being brought to the investigation. This might be clinical or process knowledge relevant to the incident but also facilitation skills (e.g. interviewing).

Example of a criterion that was not re-worded but entered subsequent Delphi Round

> Criterion 15: Investigation leads have at least 30 hours training and skills development in learning from patient safety incidents. (Wording used in round 1)

This criteria achieved a Median of 6.5, MADM of 1.5 and wording score of 62.3%. Most of the panellists comments related to issues with the fundamental meaning of the criteria, rather than the wording, with some examples below:

> *"Is 30 hours arbitrary?"*
>
> *"I'm not sure putting an hour target on this is useful. Likewise, this (to my mind) seems a low bar, in the effort to professionalise and legitimise investigation as a critical activity. A few years would be nice, with an apprenticeship model for those early on in the process (eg perhaps still in the first 500 hours!).*"
>
> *"Not sure setting a time is helpful - it's about competency, which is only partially related to exposure. Either way, 30 hours is very short - investigation needs to be professionalised, which suggests far greater hours.*"

Based on these and other comments this criteria was not re-worded but entered into round 2, at which point it achieved consensus with a median of 6, MADM of 1 and wording score of 75.5%.

Example of a criterion that was removed from Delphi

> Criterion 66: The recommendation is system-focused resulting in improvement in patient safety across different healthcare settings. (Wording used in round 2)

This criterion achieved a median of 7, MADM of 1 and wording score of 61.3% in round 1 and was re-worded and entered into round 2. In round 2 it achieved a median of 7,

MADM of 2 and wording score of 59.2%. Most of the comments related to the fundamental meaning of the criterion rather than simple adjustment to the wording. The panel questioned whether local hospital investigations can legitimately propose recommendations that span healthcare settings. They also suggested that while further reaching recommendations might be appropriate, it should not necessarily be a key indicator of a 'good recommendation', but rather dependant on the specific circumstances of an incident and investigation. Below are some examples of panellists' comments:

> *"It may not be possible to apply the recommendations in a variety of settings. It may be harmful to try to apply recommendations to other settings without fully understanding the implications."*
>
> *"Only if there is supporting evidence for such a broad scope, agree if the relevant incident covered multiple healthcare settings."*
>
> *"It makes sense, but I'm not sure how realistic it is for one investigation to improve patient safety across different healthcare settings."*
>
> *"It's written clearly; just have issues with it. It presumes that improvement may have any relevance to other settings. What may be very beneficial in the operating theater may have no bearing anywhere else. To me, this seems like the "system-focused" is prioritizing improvements across a health system rather than system-focused as in sociotechnical systems. The latter should be the primary concern in my opinion. If applicable, improvements should be shared across healthcare settings. Context matters."*
>
> *"Again, I am concerned by a systems based approach in extremis. Nearly always I would agree with this, but does it fit for all circumstances? Human agency may be the realm of improvement and changing the wider system unhelpful. This is too myopic in my view to help."*

Based on these and other comments the criterion was removed at this stage, as it was felt that panellists had concerns about the fundamental meaning of the criteria and it was unlikely to achieve consensus.

Table 5.2 84 Criteria That Would Indicate a Good Investigation Or Recommendation (Ranked by median, then MADM, then the number of sources from which the draft criteria was originally drafted in stage 1, and finally by % wording agreement) A median of 9 indicated that participants very strongly agreed that the criteria would indicate a good investigation or recommendation, 8 strongly agreed, 7 agreed, 6 slightly agreed.

Sources	Censensus achieved in	Final Median	Final MADM	Final Wording	Criteria (Final wording)
14	1	9	0	86.9	Criteria 22: Harmed patients, families, and carers are involved and engaged in the investigation, and if not, it is clear why this is the case. They have an opportunity to raise questions they want answering and are asked to provide their story. They know who to contact, what is happening when, where to find out more, and they are asked if they want to provide feedback on a draft report. (14 sources)
6	1	9	0	87.9	Criteria 10: Investigation team members are given protected time, within their usual working hours, to carry out the investigation. (6 sources)
16	1	8	1	90.8	Criteria 23: It is clear that people from a range of backgrounds are engaged in the investigation, and their perspectives taken into account. Questions from those engaged are explored and answered or responded to within the investigation. (16 sources) Backgrounds might include carers, physiotherapists, pharmacists, dietitians, doctors, nurses, ward clerks, health care assistants, administration, laboratory, maintenance, and managers. Engagement might include supporting people to identify issues, possible contributory factors, recommendations and actions. They provide feedback on the investigation analysis throughout and before final reports are published.
14	1	8	1	90.2	Criteria 40: An array of tools and approaches*, appropriate** for the investigation, are selected by investigators and justification/rationale provided. (14 sources) *examples: direct observation of practice or environments, recreating the events by “walking the process,” group meetings with involved members **Tools, approaches and frameworks have been developed for or adapted for use in healthcare. They are

					aligned with the purposes/objectives of the investigation. Investigators have training or experience in using tools.
14	1	8	1	80.6	Criteria 35: 'Human Error' is considered as a symptom of a system problem. 'Human error' is not concluded to be the 'cause' of the incident. Language does NOT directly or indirectly infer blame of individuals, teams, departments, or organisations and/or focus on human failure (i.e. the nurse failed to follow policy; the doctor lost situation awareness). (14 sources)
12	1	8	1	90.2	Criteria 63: Recommendations are demonstrably linked to the evidence and findings of the investigation and views expressed by stakeholders. (12 sources)
12	1	8	1	89.8	Criteria 77: Harmed patient or family engagement is documented (preferably in their own words), including their reflection on the completed investigation. (12 sources)
12	1	8	1	88.1	Criteria 38: Local rationality is considered. An attempt has been made to understand why decisions and actions taken by individuals involved felt right at the time, taking into account situational factors, operational pressures, and organizational norms existing at the time. People do things that are reasonable, or rational, based on their limited knowledge, goals, and understanding of the situation and their limited resources at the time. (12 sources)
11	1	8	1	88.7	Criteria 65: Recommendations do not 'blame' a suitable culprit, or assign liability, and extend beyond the behaviour and shortcomings of individuals to the wider systems factors which allowed the problems to occur. People-focused solutions like adding in another double-check, re-writing the safety procedure, creating a new safety procedure or sending out an email reminder to staff are unlikely to lead to sustained improvements, if used in isolation. (11 sources)
10	1	8	1	79	Criteria 51: The varieties of human work are identified: Work-as-Done, Work-as-Imagined , Work-as-Prescribed, Work-as-Disclosed. Analysis does not simply map out procedures and guidelines. (10 sources)
9	1	8	1	93.4	Criteria 64: Evidence external to the investigation is considered when developing recommendations. This might include research evidence or published incident investigations, "best practices" or practice guidelines that are recommended by professional organizations, trends in local incidents, findings of other local investigations, ongoing quality improvement work, complaints etc. (9 sources)

8	1	8	1	93.5	Criteria 60: Recommendations are developed collaboratively with relevant stakeholders (e.g. harmed patients/families/carers, staff, commissioners), who are able to comment/feedback on them before selection or actions are developed. (8 sources)
8	1	8	1	79	Criteria 55: Work-arounds and how and why they occur, are understood. The analysis does not stop at identifying 'noncompliance' by healthcare professionals, but identifies and describes WHY non-compliance occurrs? If memory demands are high, practitioners are likely to develop their own aiding strategics (e.g., notes, external reminders) to compensate or to simplify how they use the technological devices to reduce the need to remember so much. If there is a proliferation of displays, windows, and options, practitioners have been observed to tailor the device and their strategics to reduce the knowledge and attentional demands. (8 sources)
7	1	8	1	98.4	Criteria 58: Other sources of evidence are considered, such as relevant scientific literature, the findings of other incident investigations (local or external), or trends in incident reporting. (7 sources)
6	1	8	1	93.4	Criteria 48: There is explicit consideration of interacting system-based performance influencing factors (e.g. task complexity, technology, work procedures, workplace design, information transfer, clinical condition of patient, stress, fatigue, culture, leadership/ management, policy/regulation). (6 sources)
6	1	8	1	91.8	Criteria 4: The investigation team includes someone who has knowledge of the system or clinical area in which the incident occurred (process/domain expertise). (6 sources)
6	1	8	1	77	Criteria 1: The investigation is carried out by a team rather than an individual. (5 sources)
5	1	8	1	88.7	Criteria 46: It is clear that problems that cross organizational boundaries are identified and addressed. Different locations of care are considered, such as home, community and hospital. Harm and failures of care that patients may suffer are often due to an accumulation of problems across multiple contexts. (5 sources)

5	2	8	1	85.7	Criteria 6: The investigation team includes, or is supported by, someone with expertise in applying Human Factors and Ergonomics (HFE) in patient safety. (5 sources + panel) Further explanation: For clarity we provide a definition of HFE from the International Ergonomics Association: “Human Factors is concerned with the understanding of interactions among humans and other elements of a system. It’s the profession that applies theory, principles, data and methods to design to optimise human wellbeing and overall system performance. Practitioners contribute to the design and evaluation of tasks, jobs, products, environments and systems to make them compatible with the needs, abilities and limitations of people.”
5	1	8	1	85.5	Criteria 44: It is clear that information is actively collected from a wide range of sources, such as: · medical records (e.g. nursing, medical, community, social workers, GP) and all correspondence, including internal communications · both electronic and written records · documentation and forms related to the incident (e.g. relevant protocols and procedures) · immediate statements, or observations · physical evidence (e.g. ward or incident site layout/schematics) · secure equipment involved in the incident (e.g. a cardiotocography machine or medication pump implicated in a case) · information about relevant conditions affecting the event (e.g. staff rota, availability of trained staff) · results of interviews or collation of statements from persons involved in the incident early, so that memorable information is not lost. · biomedical equipment, IV solutions, medications, packaging, and garments. · Photographs of the items and workspace are often helpful.
5	1	8	1	85.2	Criteria 24: All knowledge should be valued equally; and all voices, should be heard equally. Conflicting views are not ignored, and no single perspective or professional group dominates. (5 sources)
4	1	8	1	86.9	Criteria 78: The written report is clear, easy to read and anonymised. It is written in plain English, using inclusive language. It is written to ‘inform rather than impress’. (4 sources)

4	2	8	1	81.3	Criteria 37: There is a focus on understanding what happened and why and not on who should have done (or not done) what. (4 sources)
4	2	8	1	79.6	Criteria 2: The investigation team is made up of people from different professional and non-professional backgrounds to bring a range of perspectives and experiences relevant to the incident. (4 sources + panel) Further explanation: By 'different backgrounds' we mean nurses, doctors, HCAs, porters, administrative staff, volunteers etc. People who are capable of providing different perspectives, which may include non-professional staff as well as professionals, depending on the type of investigation. This is to increase the chance of diverse perspectives being captured and a range of skills and expertise being brought to the investigation. This might be clinical or process knowledge relevant to the incident but also facilitation skills (e.g. interviewing).
3	1	8	1	96.8	Criteria 53: There is evidence of reflection on the workability of the underlying care process. The Investigation looks closely not only at the reasons for departures from standard procedures but at the standards and procedures themselves. The feasibility and workability of current standards and practices is considered and whether these need to be adjusted. (3 sources)
3	1	8	1	95.2	Criteria 76: The investigation report is shared with affected patients, relatives, carers, and staff before publication. This is while it is still in draft and there is a realistic possibility that their suggestions may lead to amendments. (3 sources)
3	1	8	1	91.8	Criteria 28: Those staff involved in, and affected by, the investigation are given information about what to expect, who is on the investigation team, where to find out more, and who to contact. Information is provided in easy to understand language. (3 sources)
3	1	8	1	90.2	Criteria 83: The report is available to stakeholders and staff. (3 sources)
3	1	8	1	78.7	Criteria 11: The investigation lead is skilled in creating psychological safety for the team and those engaged. (3 sources)
2	1	8	1	95.2	Criteria 52: Disagreements and conflicting information are not avoided or reduced. Disagreements or discrepancies are clearly identified. (2 sources)

2	1	8	1	95.1	Criteria 86: The rationale for the recommendations is clear. (2 sources)
2	1	8	1	91.7	Criteria 29: Those involved in, and affected by, the investigation are informed of sources of support and provided with access to counselling. (2 sources)
2	1	8	1	85.2	Criteria 14: Investigations are led by experts in safety investigation, who are able to draw on skills, experience and expertise in patient safety, improvement science, human factors, healthcare provision and clinical services. (2 sources)
1	1	8	1	93.4	Criteria 62: Time and space is given to problem-solving in relation to recommendations. This might be a dedicated meeting focusing on the formulation of recommendations considering quality, feasibility and effectiveness, as perceived by staff. (1 source)
1	1	8	1	91.8	Criteria 61: People with human factors expertise are involved in developing recommendations. (1 source)
1	1	8	1	90.3	Criteria 92: The general lessons and findings are disseminated within, and where applicable, outside the organization to prevent harm recurrence. (1 source)
1	1	8	1	90.2	Criteria 16: Investigation leads undertake continuous professional development in incident investigation skills and knowledge, and network with other leads at least annually to build and maintain their expertise. (1 source)
1	1	8	1	86.4	Criteria 33: Interviews are held as soon as possible after the event. (1 source)
1	2	8	1	83.7	Criteria 27: Staff are told about their involvement in an investigation in a timely, sensitive, compassionate and supportive manner.(1 source)
1	2	8	1	83	Criteria 13: Investigation leads are open to the views and expertise of others, which is evident in the investigation findings and final report (1 source)
1	1	8	1	82	Criteria 25: It is clear that staff are given protected time to engage/take-part in the investigation. (1 source)
1	2	8	1	81.6	Criteria 31: Investigators should be able to establish a psychologically safe environment where people can express their feelings', and these are acknowledged and respected by others. (1 source)
1	1	8	1	80.6	Criteria 36: There is a clear distinction between the 'honest mistakes' of well-intentioned healthcare workers, where punitive responses are neither warranted nor helpful; and the rare acts that involve reckless neglect or mistreatment. (1 source)
3	1	7.5	0.5	88.5	Criteria 54: Complexity is not over-simplified. Complexity has been uncovered in interactions between system components across time. Accidents result from flawed interactions between components of a system, not just failures in individual components. (3 sources)

2	1	7.5	0.5	86.9	Criteria 74: Recommendations are accompanied by a plan to monitor progress over time. (2 sources)
10	1	7	1	87.1	Criteria 90: Measures are selected to monitor recommendations and subsequent actions, such as implementation, reasons not implemented, associated costs, effectiveness, unintended outcomes (negative or positive), similar incident recurrence. (10 sources) In relation to the measures - the following are clearly defined: • a description of the measure • units of measurement (and any formula for its calculation) • how data will be collected • measurement frequency • how the data will be visualised/displayed • reporting intervals.
6	1	7	1	86.9	Criteria 50: There is a description of how issues are normally solved. How is clinical care usually achieved safely and how are issues detected, anticipated and recovered from? (6 sources)
5	2	7	1	87.8	Criteria 7: The investigation team includes someone who has expertise in the investigation of incidents in healthcare. (5 sources + panel) Further explanation: By this we mean they have training and experience in applying different methodologies and appropriate analytical approaches to perform investigations. We feel that defining the exact requirements for expertise is beyond the scope of this Delphi.
5	2	7	1	85.7	Criteria 9: No member of the investigation team was directly involved in the incident being investigated. (5 sources)
5	1	7	1	83.6	Criteria 71: The potential resource and financial burden of actioning a recommendation is considered. Cost-benefit has been considered. (5 sources)
4	1	7	1	93.5	Criteria 91: There is a plan to involve, or at least inform, people affected by an incident in the follow-up of investigations, recommendations and subsequent actions. (4 sources)
4	1	7	1	90.3	Criteria 73: The potential negative impact of recommendations is considered and will be monitored. Proactive risk assessments of recommendations and actions is done to understand their wider impact on systems. (4 sources)

4	1	7	1	85.5	Criteria 59: Recommendations and actions are clearly separated. Recommendations do not skip directly to the action/solution. During the process of investigation, and working with stakeholders, recommendations should be developed. Actions may be suggested but there is separation of the investigation and implementation teams. (4 sources)
4	1	7	1	85.5	Criteria 88: A risk management group, clinical governance team, or organisational board are made responsible for implementing and tracking recommendations. (4 sources)
4	2	7	1	83.3	Criteria 8: The investigation team includes, or is supported by someone, who has expertise in understanding how biases can affect investigators and investigations. (4 sources + panel) Further explanation: A bias is a strong feeling in favour of or against one group of people, or one side in an argument, often not based on fair judgement.* People are often not aware that their judgements are affected by biases. There are probably more than 180 biases that affect our thinking; one example is hindsight bias which is the tendency for people to think something was more likely to happen, once they know the outcome, sometimes referred to as the 'Knew-it-all-along' effect. *https://www.oxfordlearnersdictionaries.com/definition/english/bias_1
4	1	7	1	82	Criteria 67: Recommendations should be developed which target different levels or areas of the system. An example of this might be recommendations that target leadership level, separately or in addition to middle managers and those who provide services/deliver care. (4 sources)
3	1	7	1	88.3	Criteria 41: The incident(s) 'story' is described in detail as a narrative that shows how the events unfolded, over time. (3 sources)
3	1	7	1	83.3	Criteria 17: A flexible timescale is agreed with stakeholders (investigators, commissioners/leadership, patients/families/carers, and staff) and adjustments to this are agreed by all. (3 sources)
3	1	7	1	80.3	Criteria 43: It is clear that when explanations are sought, a number of alternatives are considered during the analysis. "Outside" perspectives are sought. Explanations are revised as new information/perspectives are obtained. (3 sources)
3	2	7	1	77.6	Criteria 47: It is clear that the patient journey is fully explored. Analysis may initially focus on the time period where problems were most apparent, but that should not prevent investigators examining other parts of the patient journey if it is considered useful. (3 sources)

2	2	7	1	85.7	Criteria 45: The investigation findings have been constructively critiqued and challenged by subject matter experts, independent of the investigation. (2 sources)
2	1	7	1	85.5	Criteria 70: It is described how the recommendation will reduce the risk or limit the consequences of a similar incident. (2 sources)
2	2	7	1	81.6	Criteria 49: Factors affecting different parts of the patient journey or clinical pathway are examined.
1	1	7	1	98.3	Criteria 80: The methods for analysis are specified. (1 source)
1	1	7	1	95.2	Criteria 87: Reports are produced using an organisational standardised template. (1 source)
1	1	7	1	91.9	Criteria 34: Staff, not involved in the incident, but with knowledge about the care process contribute to the investigation. (1 source)
1	1	7	1	91.7	Criteria 69: The use of tools or methods used to develop recommendations are explained and justified. (1 source)
1	2	7	1	87.8	Criteria 19: Management support is clear and visible.(1 source + panel) Further explanation: A named manager or managers are identified who can support the investigator and investigation, helping deal with challenges or barriers. This might be ensuring necessary resources or access to information, such as patient records. Managers may be able to help with ensuring participation of individuals by, for example, releasing staff from clinical duties, to be involved.
1	1	7	1	83.9	Criteria 84: The aftercare for the patient/relatives/carers is described. (1 source)
1	1	7	1	83.3	Criteria 30: Informal interviews, with those involved in, and affected by, the investigation should be conducted one person at a time so that individual perspectives about the incident are well understood. (1 source)
1	1	7	1	80.6	Criteria 75: Recommendations are produced to meet the purposes of the investigation. (1 source)
1	1	7	1	80.6	Criteria 82: The report states who is responsible for any further recommendations. (1 source)
1	1	7	1	80.3	Criteria 85: The aftercare for the staff involved is described. (1 source)

1	2	7	1	77.6	Criteria 79: The description of the event gives a complete picture of the relevant elements of the work systems and patient journey. (1 source) Further explanation: Frameworks such as the System Engineering Initiative for Patient Safety (SEIPS) and the Yorkshire Contributory Factors Framework (YCFF) can be useful in identifying relevant elements.
1	1	7	1	77.4	Criteria 3: The investigation team includes a patient representative, who is unrelated to the incident being investigated. (1 source)
1	1	7	1	76.7	Criteria 26: Engaged stakeholders are allowed to bring a friend, family member or advocate of their choice with them to any meeting, that is part of the investigation, they are involved in. (1 source)
1	1	7	1	75.4	Criteria 39: The investigation report is proofread by subject matter experts before publication. (1 source)
4	1	6	1	77.4	Criteria 89: Follow-up is assigned to an individual not a group or committee. The implementation and efficacy of all recommendations are monitored, and a named individual identified with responsibility for this. (4 sources)
3	1	6	1	65	Criteria 12: Investigations leads are senior members of staff. (3 sources)
2	1	6	1	66.1	Criteria 57: Diagramming is used as a helpful exercise in understanding the relationship between contributing factors. The Tree and Constellation Diagrams are two potential tools for diagramming. (2 sources)
2	1	6	1	54.1	Criteria 5: The investigation team is informed in change management. (2 sources)
1	2	6	1	75.5	Criteria 15: Investigation leads have at least 30 hours training and skills development in learning from patient safety incidents. (1 source)
1	1	6	1	61.3	Criteria 68: The recommendation fundamentally alters how things get done within the targeted context. It reflects a paradigm shift in how the problem is perceived and the strategies used to address it (e.g. viewing and doing things differently). (1 source)
1	2	3	1	66	Criteria 18: The investigation is completed within 45 days of declaration that an investigation is needed. (1 source)

Table 5.3. 8 Criteria that failed to achieve consensus by the end of round 3 or were removed from the Delphi process. (Ranked by median, then MADM, then the number of sources from which the draft criteria was originally drafted in stage 1, and finally by % wording agreement)

Sources	Final Median	Final MADM	Final Wording	Criteria	Rationale
4	8	1	67.2	Criteria 21: The purpose of the investigation is to identifying opportunities to improve quality and safety, and reduce risk. (4 sources)	Criteria 20 and 21 were combined into a single criteria following panel comments at round 1. After review of round 2 panel comments the steering group agreed that these statements represented purpose statements for an investigation, rather than a criteria to judge the quality of an investigation.
4	7	1	61.7	Criteria 20: The purpose of the investigation is to learn about the organization, how it constrains or supports the people to deliver high quality care. (4 sources)	
1	7	2	59.2	Criteria 66: The recommendation is system-focused resulting in improvement in patient safety across different healthcare settings.	The rationale for removal of this criteria is discussed in the body of the results.
1	6.5	1.5	77	Criteria 56: Contributory factors are not described in terms of their perceived importance in relation to the incident.	Failed to achieve consensus after round 3.
1	5	1.5	48.4	Criteria 81: The report allows you to answer these criteria.	This criteria achieved a low median and on review of panel comments the steering group felt this would not be improved with re-wording.
5	5	2	73.8	Criteria 72: Recommendations without sufficient evidence, or cost-benefit are not actioned. (5 sources)	Following review of scores and panel comments it was felt the meaning of this

					criteria was very similar to criteria 71 and therefore they were amalgamated.
1	5	2	71	Criteria 32: The board of directors provide their perspective on the analysis, conclusions and recommendations in the report.	Failed to achieve consensus after round 3.
1	4	2	53.1	Criteria 42: Care delivery problems (CDPs) are identified. These may be slips, such as picking up the wrong drug, lapses of judgement, forgetting to carry out basic observations or, rarely, deliberate departures from safe practices, procedures or standard. Ensure that all CDPs are specific actions or omissions on the part of the staff, rather than more general observations on the quality of care, which should be recorded elsewhere. It is easy, for example, to put down 'problems with teamwork' as a CDP which may be a correct description of the team, but should be recorded as a contributory factor as it was likely that problems with teamwork influenced the CDP. (1 source)	This criteria failed to achieve consensus in round 1. Most panel comments related to issues with the fundamental meaning of the criteria rather than wording. It was not re-worded but entered into round 2. It failed to achieve consensus in round 2, but again most panel comments related to the issues with the fundamental meaning. The steering group agreed to remove this criteria after round 2 due to the panel comments and low scores, in order to reduce the burden on panellists and reduce drop out.

5.4 Discussion

The Delphi study resulted in 83 criteria that experts agreed would indicate a good investigation or recommendation. Criteria were organised into six domains: 1) planning investigations and selecting investigators, 2) engaging people in the investigation, 3) using a systems approach to gathering data, information and evidence , 4) developing recommendations, 5) documenting the investigation, and 6) follow-up of the investigation and recommendations. To the authors' knowledge, this is the first empirical research study to develop such criteria providing investigators, policy makers and researchers with guidance that was previously lacking.[15,254]

The findings of this are important as they have the potential to address the lack of clarity around what constitutes a good investigation or recommendation.[254] These criteria could be utilised by those investigating incidents or with oversight responsibility to perform a gap analysis and identify opportunities to improve practice, processes and infrastructure. The criteria could ultimately be used to improve the impact of investigations in reducing risk and improving patient safety.

It will be useful to consider how these criteria align with the wider literature. The five most highly rated criteria, representing those criteria that were felt by the expert panel to be most important in identifying a good investigation or recommendation, are examined here in more detail.

The highest ranking of the criteria (number 22, table 5.2) referred to the involvement of harmed patients, families and carers in investigations. This is somewhat surprising, as this has been done poorly if at all to date,[48,333] but that this is rated so highly may reflect a more recent drive for change in policy, evidence and from harmed patients. While there may be challenges to effectively involving this group, evidence suggests they can provide valuable and unique perspectives,[334-336] and their participation in safety activities is vital.[337]

The potentially harmful effects of the investigation process itself, on patients, families and staff, are well known,[48,338-344] and so must be improved to help restoration and healing after incidents.[256] With the increasing awareness of the potential benefits of patient and family involvement and need to avoid compound harm it is maybe not surprising that this criteria has ranked so highly.

It is not surprising that protected time is considered important to the quality of an investigation(criteria 10, table 5.2), and that improvements in patient safety and quality of care are unlikely to be achieved without dedicated resources and the release of staff from other duties, such as direct patient care.[345] Health services are under considerable pressure,[346:p.2024] and it is not difficult to see how many organisations may struggle to prioritise time and resource for lengthy investigations. It is therefore important to consider a range of potential responses to incidents, such as aggregated investigations or after action reviews within a system of safety learning; in other words using 'safety time' more wisely.[82] It is encouraging to see the introduction of a new framework for responding to incidents in the English NHS, which promotes a proportionate response to incidents.[105]

The third highest rated criteria (23, table 5.2) refers to the engagement of a wide range of stakeholders or perspectives. Healthcare is a highly complex dynamic sociotechnical system, in contrast to other industries, there is a high level of personal interactions, contingencies that cannot be fully anticipated and different, sometimes conflicting, perspectives and goals of providers and users.[108,109] In order for investigators to fully appreciate the complex systems in which incidents occur, and the potential areas for improvement, effective engagement with the right stakeholders is vital.[345,347,348]

The fourth highest rated criteria (40, table 5.2) relates to the need for explicit rationale for the use of tools and approaches appropriate for the analysis of the incident to be investigated. Despite it's continued wide-spread use in healthcare,[254] the way in which Root Cause Analysis (RCA) has been used has come under significant scrutiny.[12] While a number of problems have been

identified,[12] two that seem particularly relevant to this criteria are its predominance over alternative methods[254] and lack of adaptation for the healthcare context or empirical evaluation.[12,213] While some work has been undertaken to develop evidence-based approaches, adapted for the healthcare context, for analysis and investigation,[82,179-181,236] further work is needed.

The fifth criteria (35, table 5.2) relates to the continued predominance of investigations and subsequent recommendations to focus on improving individuals' behaviour and practice, rather than the wider system deficiencies that contribute to incidents.[10,145,238-245,254,255,349] A number of potential causes for this have been identified such as lack of investigatory expertise,[10,156,180] the limited guidance which might itself encourage a focus on individuals,[12,144] and even investigator bias.[198,272,349] Not only is an inappropriate focus on individuals within investigations and recommendations likely to result in missed opportunities to identify and improve other system deficiencies, it will perpetuate the creation of second victims and contribute to disenfranchisement and burnout within the healthcare workforce.[350] Criteria 35 is intended to encourage investigations to look beyond the active failures of generally well-meaning healthcare staff, to the underlying systemic deficiencies that provide they environment and context for incidents.

5.5 Strengths and Limitations

Utilising both peer-reviewed and grey literature in combination with expert opinion likely strengthened the validity of these criteria, which is suggested by their alignment with the broader literature above. The healthcare-focused expert panel are likely to have increased the likelihood that the criteria are adapted for the healthcare context, increasing their potential relevance and usability.[213,237]

The average wording score across the criteria was 86.5% (SD 5.6) indicating that while the wording of the criteria was considered acceptable, it could be improved.

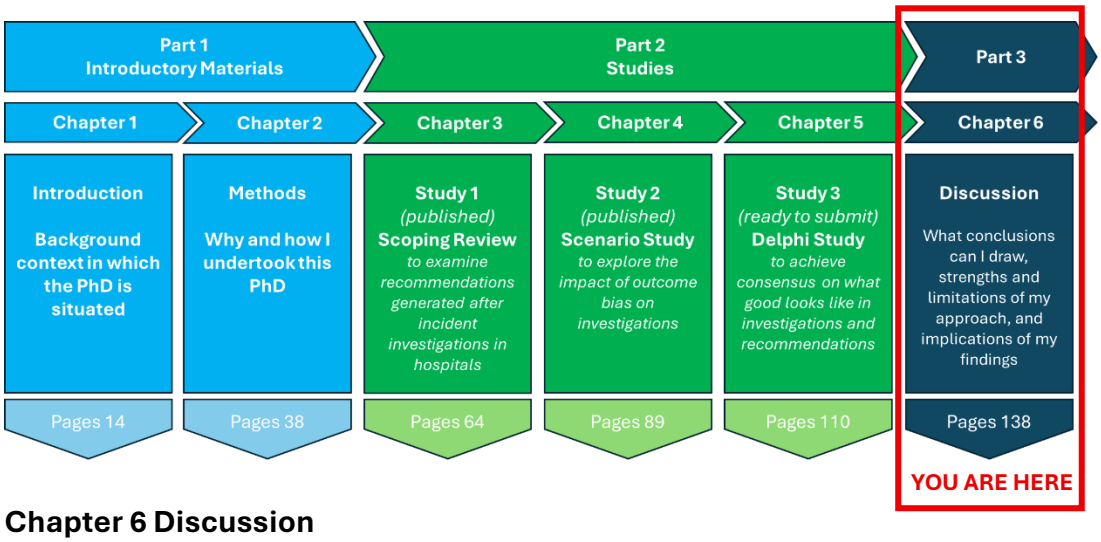
It is likely that further work is needed to establish exactly how these criteria can be used in practice, and there are likely to be limitations to this, in their current form. One potential limitation for the use of these criteria is the availability of evidence or data from investigations or their reports; it may not be possible to judge an investigation against criteria when information is not available. As an example, criteria 10 states that "Investigation team members are given protected time, within their usual working hours, to carry out the investigation.". It is unlikely that this would be documented within a single investigation report, but might be written within a policy document. The criteria as they are currently presented, are potentially aspirational, and it will be important for further work to consider how they are evidenced and applied in practice.

It is also likely that the criteria are incomplete; i.e. that there are other indicators of a good investigation or recommendation, that have not been identified during this Delphi. For instance, despite the increasing awareness of inequity within patient safety, and the potential impact of explicit or implicit racism, sexism, or homophobia,[351,352] there are no criteria that relate to ensuring investigations do not contribute to or further exacerbate these problems.

5.6 Conclusion

Improving the quality of investigations and recommendations is of paramount importance. Firstly because of the growing criticism that investigations and recommendations, in their current form, are failing to improve safety.[9-13,254] Secondly the increasing realisation that investigations have the potential to cause harm to patients, families and staff.[256] Thirdly that recommendations may not only be ineffective but may contribute to safety clutter or indeed have negative impacts on safety.[277,278] Finally investigations and subsequent recommendations represent a significant resource burden.[254] If the resource expenditure is to be justified then the quality of investigations and recommendations most be more explicitly demonstratable and opportunities for improvement identified and acted upon. The criteria from this Delphi provide an empirically developed and healthcare contextualised guide to achieve this.

Part 3 Discussion and Conclusions



Chapter 6 Discussion

Chapters three, four and five, each containing discussion sections, have specifically addressed the aims and objectives of the PhD outlined in the introduction (section 2.4). This thesis discussion will bring these studies together in considering how this PhD has added to existing literature around recommendations generation and the implications for practice, policy and future research.

6.1 Overview of studies

The aim of this PhD and thesis was to explore how to produce effective recommendations following investigations in hospitals. Chapter 3 presented a scoping review that updating and expanded on a previous review by Card and colleagues.[10] This update demonstrated the continued tendency for patient safety investigations to produce recommendations that focussed on improving individuals' behaviour and practice, rather than the wider system deficiencies that contributed to incidents. It highlighted a lack of transparency about how recommendations are generated and confusion, or lack of consensus, about how they should be judged for effectiveness or quality. Overall chapter 3 suggested that many questions still remained about how to generated effective recommendations. The individual focus of recommendations and question of recommendation quality were further examined in chapters 4 and 5.

The tendency for individual focussed recommendations was explored further in chapter 4, the scenario study, which demonstrated the impact of outcome bias on investigators. More punitive recommendations were selected when patient outcomes were worse following a patient safety incident. Furthermore, worse patient outcomes were associated with greater responsibility assigned to healthcare staff involved in the incident and a stronger motivation to investigate. Interestingly participants motivation to investigate was high, irrespective of patient harm, but further increased as patient harm increased. All incidents were considered by all participants to be highly avoidable irrespective of patient outcome. Chapter 4 highlighted important differences between the public, staff and experts, in their responses to an incident and investigation. Expertise in patient safety reduced, but did not eliminate, the impact of outcome bias on judgements of staff responsibility and the selection of punitive recommendations. Differences in responses between those experts from a clinical background versus those from a non-clinical background suggest that the origin and types of expertise and experiences are likely to be important in determining the effectiveness of investigations and recommendations. Chapter 4 empirically demonstrated that those with patient safety expertise were less likely to select individual-focused recommendations, than staff or the public. This highlights the importance of investigator expertise in achieving effective investigations and recommendations.

The Delphi study, presented in chapter 5, was designed to address the lack of consensus on how to judge recommendation quality, highlighted in the chapter 3. This was the first empirical research study to develop criteria to judge the quality of investigations and recommendations specific to the healthcare context. Despite the lack of agreement, in this area demonstrated in chapter 3 and potentially conflicting views of stakeholders highlighted in chapter 4, consensus was achieved for 84 criteria. The criteria were organised into six domains: 1) planning investigations and selecting investigators, 2) engaging people in the investigation, 3) using a systems approach to gathering data,

information and evidence , 4) developing recommendations, 5) documenting the investigation, and 6) follow-up of the investigation and recommendations.

Both the demonstrated impact of outcome bias on investigations and recommendations and quality criteria represent important areas of novelty within this PhD and thesis. The scenario study represents, to the PhD candidates' best knowledge, the first empirical study of outcome bias within the context of healthcare incident investigations. In the following section these two issues will be explored further within the wider context of literature, policy and practice.

6.2 Key findings

The following sections will present some key findings from this thesis and the studies within this PhD. These will be discussed in the broader context of literature and policy.

6.2.1 Individual-focus of recommendations

Despite over two decades of discussion of safety theory in healthcare policy and practice, particularly the influential work of James Reason, the problem of individual focus remains. Policy attempts to guard against the individual focus observed within investigations, but human tendencies such as outcome bias illustrated in chapter 4, make this challenging.

The scoping review in Chapter 3 demonstrated the persistent widespread individual-focus of recommendations generated in hospitals following investigations. The scenario study presented in Chapter 4 provided evidence for outcome bias as a potential mechanism for this. Taken together, these findings suggest that simply telling people that they should not focus on individuals within investigations and in the generation of recommendations, within policy, guidance or training, is unlikely to be sufficient to overcome the impact of unconscious cognitive bias.

Outcome bias may be one reason the patient safety movement has struggled to shift the focus of investigations from individuals to take into account other elements of the system.[10,100,254,353] Shifting this focus has been and must continue to be priority for a number of reasons. Firstly, being involved in an incident in itself can be traumatising for staff,[338] but the retributive or punitive approach of investigations can further compound this harm.[48,339-344] Of particular relevance in the context of a current global workforce crisis,[350] is that staff leaving the their jobs after incidents and investigations contribute to high staff turnover.[354-356]

Secondly, the focus on individuals and humans as the 'problem' has detracted from the identification of wider systemic issues and a design approach to healthcare safety.[357] James Reason used the analogy of swatting mosquitoes rather than draining the swamp, suggesting that "*active failures are like mosquitoes. They can be swatted one by one, but they still keep coming. The best remedies are to create more effective defences and to drain the swamps in which they breed. The swamps, in this case, are the ever-present latent conditions.*".[87:p769] In other words individuals who err can be removed, punished, trained or reminded, but improving safety requires a focus on how the system is designed around them to optimise their performance.87,357 This is similar in principle to Johnston's substitution test.[358] Johnston suggested that when considering the unsafe acts of individuals, it is important to consider whether another similarly qualified or experienced individual might err in the same way, particularly if placed in similar circumstances and conditions.[358] If this is the case then a focus on individuals would potentiality ignore wider system factors.

There are however criticisms of shifting the gaze of investigations, in an attempt avoid blame, could however be problematic for two reasons. Firstly, this might ignore those staff who are knowingly negligent or do intentionally cause harm to patients.[359] Secondly, and potentially more importantly, humans represent a central component within healthcare systems, and ignoring their contribution to overall system performance, within investigations, would result in incomplete

analysis and potentially ineffective recommendations. Incident analysis needs to achieve a detailed understanding how unavoidably fallible humans err as well as succeed within the systems they work in order to design or re-design systems to improve safety. While investigations would benefit from a broader systems approach, the human contribution should not be ignored and in fact explored and researched further.[357] This greater understanding of human performance within the healthcare context will allow effective system design to optimise human performance, wellbeing and overall outcomes. For this to be done effectively it might be that the assignment of blame and individual responsibility need to be de-coupled from the investigation process. The scenario study in Chapter 4 demonstrated that responsibility and blame are potentially associated with unconscious tendencies and therefore will prove challenging to mitigate. Another final consideration in relation to individual responsibility is the conflicting views of the public, staff and experts. Each of these groups may have different perspectives, judgements and expectations shaped by their experiences. These will need to be considered and managed.

6.2.2 Assessing the quality of recommendations

Improving the quality of investigations and recommendations is of paramount importance for a number of reasons. Firstly because of the growing criticism that investigations and recommendations, in their current form, are failing to improve safety.[9-13,254] Secondly the increasing realisation that investigations have the potential to cause harm to patients, families and staff.[256] Thirdly that recommendations may not only be ineffective but may contribute to safety clutter or indeed have negative impacts on safety.[277,278] Fourthly the number of recommendations proposed at local and national levels has increased exponentially with the accumulation of public enquires (e.g., Kirkup 2015; Ockenden 2022, Infected Blood Inquiries 2022) and establishment of national-level independent investigatory bodies such as the HSSIB.[360] Investigations and subsequent recommendations represent a significant resource burden, with scant evidence they are implemented.[254] If the resource expenditure is to be

justified then the quality of investigations and recommendations most be more explicitly demonstratable and opportunities for improvement identified and acted upon.

The scoping review in Chapter 3 demonstrated a great deal of confusion regarding the judgement of recommendation quality. Despite its importance, reporting of outcome measures relating to investigations and recommendations in the literature was scant. Instead of evaluating effectiveness of recommendations using outcome measures there was a preference within the literature for predictive or presumptive quality measures such as the action hierarchy(AH). While widely used, in academia and policy, the adoption and iterative development of the AH within healthcare has been criticised as lacking empirical basis and adaption for the context of healthcare.[10,100,101,148,182,183,186,213,234-237,251,254,255] A challenging issue in this area was what constitutes a good quality recommendation, and the Delphi study provided an empirical step in moving forward from the confusion and disagreement highlighted in the scoping review. While the design of the Delphi study was discussed in more detail in chapter 5, one important reason for developing criteria for the investigation *and* recommendations was to allow future analysis of the linkage between these steps. The output of this study was strong agreement on seven criteria directly relating to recommendation quality (criteria 60-65, and 86). A number of the criteria will be considered in more detail here as they relate to the broader aims of this PhD and thesis.

The highest rated recommendation criterion (criterion 63) stated recommendations should be demonstrably linked to the evidence and findings of the investigation as well as views of stakeholders. This supports the original Delphi aim to identify criteria to judge the quality of both recommendations and preceding investigations as it is likely that these are inextricably linked. Recommendation generation is a step in the investigation process informed by the findings of the investigation. It could be argued that the recommendations can only be as effective as the preceding investigation.[329] This concept would

benefit from further analysis, for which the Delphi criteria may provide an empirical basis.

The second most highly rated recommendation criterion (criterion 65) related to avoiding the blame of a 'suitable' culprit or assignment of liability and that they should extend beyond the behaviour and shortcomings of individuals to wider systems factors. This is significant when considering the two other studies in this PhD. The scoping review highlighted the reality that in hospital practice there is a tendency for individual-focussed recommendations, despite consensus among experts in the Delphi that this should be avoided. Within the scenario study (Chapter 4), irrespective of patient outcome severity, 836 punitive or individual-focused recommendations were selected, representing 34% of the total selected recommendations. Simply put, this finding suggests that whatever the outcome, people continue a significant proportion of recommendations on those individuals involved in an incident. It is important to note that participants were asked to ignore any potential perceived resource implications or implementation challenges when selecting recommendations, a reported barrier to the selection of recommendations in practice. An even greater disconnect is that even experts appeared to select more punitive recommendations when the patient outcome was most severe. This may indicate both the strength and unconscious nature of outcome bias and its impact on decision making in the context of healthcare investigations and recommendation selection.

6.2.3 Bias within the investigation life cycle

This bias has been demonstrated in studies from other domains and for over 30 years suggesting it's a pervasive and persistent phenomenon.[190,191] Outcome bias has the potential to influence judgements at a number of different points throughout the investigation process, thus having a cumulative and therefore greater impact on the investigation and recommendation generation process.[349] These points represent distinct phases of the investigation from deciding whether an investigation is to be carried out, to the focus of analysis (individual responsibility) and finally where improvement efforts are focused

(recommendations). When a patient comes to greater harm an incident is more likely to be investigated, individual responsibility may be given greater attention within the analysis and ultimately the generated recommendations may inappropriately focus on individuals.

Figure 6.1 below illustrates the potential cumulative impact of outcome bias during the course of an investigation, expanding on findings of the scenario study with three other considerations.

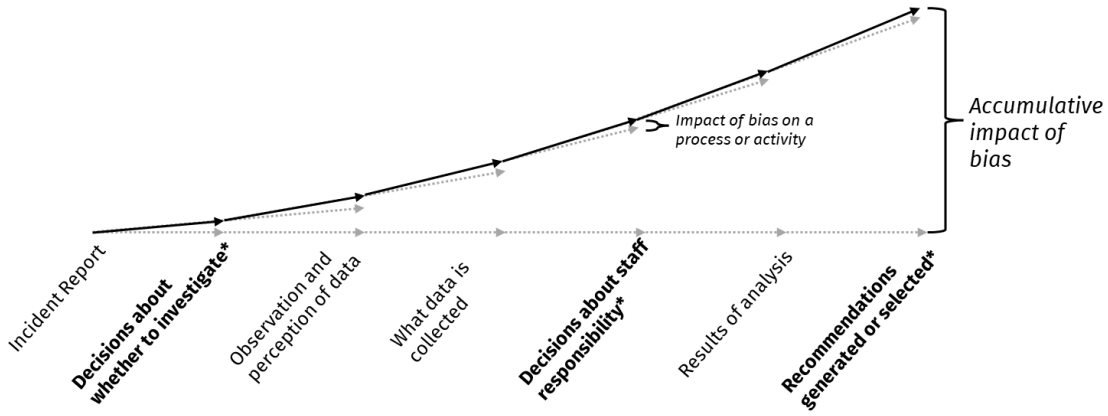
Firstly, although explored in non-healthcare domains, other points during investigations have been identified as susceptible to bias, such as the perception and observation of data, data collection and analysis.[361] It is reasonable to hypothesise therefore that outcome bias will impact on more elements of the investigation process than investigated within the scenario study reported here.

Secondly, outcome bias represents only one of possibly 180 or more biases that have been identified, outside of the healthcare incident investigation domain, as effecting decision making in a range of professional domains.[320] For example, investigators have been demonstrated to attribute greater individual responsibility for incidents than the evidence would suggest, irrespective out outcome bias.[362-364] There's also a demonstrated tendency to consider people's actions as intentional and overly impacted by dispositional qualities such as inattentiveness or carelessness.[365] Often people consider incidents, in hindsight, as more predictable than they actually were,[366] correlating with the high levels of avoidability judgements within this PhD's scenario study.[349] This perceived foreseeability of the incident may lead investigators to focus on the actions of individuals, poor decisions or missed opportunities rather than wider system issues.[271,367] Importantly, a number of biases are likely to result in investigators focusing on human error, such as the tendencies to conclude that individuals' actions are intentional or the result of carelessness or inattentiveness.[365] These are just a few examples of cognitive biases that might have an cumulative impact on the investigation process.

Thirdly, investigations in healthcare involve a lead investigator interacting with a number of other people, whether within an investigation team, interviewing patients, families and staff, or presenting to oversight committees. The scenario study examined outcome bias in the context of an individual investigator, rather than taking into the account the impact of these interactions. Teams or groups of people may themselves introduce further biasing issues such as "group think",[284] and shared information bias[285,286] which may result in other forms of bias affecting the gathering and interpreting of information, as well as the analysis and conclusions.

Taken together, these issues suggest that investigations and those investigating may be susceptible to a range of cognitive biases occurring at different time points, potentially impacting the findings and subsequent recommendations.

Figure 6.1 Accumulative impact of bias during the investigation process



The key findings of this PhD have been explored within the wider context of literature, policy and practice. In the following section the implications and potential applications of these findings will be discussed.

6.2.4 What does this mean for the practice of recommendation generation

There are important limitations of the focus of this PhD and thesis in ultimately improving patient safety. Recommendations themselves represent only one

component in the realisation of safety improvements following the identification of patient safety incidents and subsequent investigations. The 'perfect' recommendation may not lead to any improvement in safety, and it is important to briefly consider the expectations of recommendations and wider context in which they are generated.

Firstly, there is an expectation that recommendations should be implemented, and actions developed; in other words that recommendations should trigger change. In fact the scenario study highlighted the preference to select recommendations than not. Recent policy change in England has encouraged a proportionate response to incidents in that patient safety teams in hospitals should make decisions about the kinds of investigation approaches to use balancing opportunity for improvement with resource usage. It could be argued that a similarly proportionate response is needed in response to recommendations. A comparison to evidence-based medicine (EBM) may be useful to illustrate the argument. EBM uses high-quality clinical research to guide decisions around patient care. An important principle is that changes to care are only advised based on sufficient bodies of research that is judged of high quality. If a single poorly designed study with a small number of participants concluded that a new medication was effective for treating a particular condition it is unlikely that the medical community would make changes to their practice. It could be argued that individual incident investigations are analogous to the small study. Taking an EBM approach to incident investigations then it would be suggested that the recommendations and individual investigation need to be evaluated for quality and set in the broader context of the organisations patient safety evidence before deciding whether to action them. This is not an entirely new suggestion, as others have encouraged the potential value in aggregated and thematic reviews of incidents and investigations as a way to improve the potential for safety improvement.[368] If a proportionate approach to recommendation actioning was taken, there may be situations in which recommendation from individual investigations are not

actioned. The scenario study and broader literature suggests that this may be challenging as organisations need to provide patients and families with answers and staff and commissioners or regulators with assurance of safety. [159]

Secondly, it is important to consider the wider context within which recommendations are generated. Recommendations do not exist in isolation, and it is important to consider this when evaluating their effectiveness. As discussed earlier, recommendations follow on from investigations and before that, reported incidents. Within this thesis the potential unreliability with incident reporting has been highlighted. Investigations are only likely only to occur as a result of unsafe care that is identified by reporting, potentially ignoring other opportunities for learning. It is likely that the focus of the investigation is open to bias. The recommendations are likely to be largely informed by the findings of the investigation and therefore the quality of this is likely to be an important influence. There are a large number of other external factors that might influence the generation and selection of recommendations. One important factor is the potential resource implication of actioning a recommendation. The scenario study asked participants to ignore potential cost or resource implication of recommendations. While this was important in the design of the study, in examining the impact of outcome bias, it is clearly an unrealistic suggestion. In reality the potential or perceived resource implications of actioning recommendations has an impact on which recommendation are generated. As resources within healthcare become even more stretched, it is likely that this is going to have an increasing impact. Overall the argument here is that while recommendations are an important step in the investigation process they sit within a larger process with many influences. Realising the potential of investigations to improve safety will require an approach that take into account this wider context.

6.3 Implications for Practice

This PhD has important implications for how recommendations are generated and investigations are carried out in hospitals. It also has implications for independent or national organisations such as the Health Services Safety

Investigations Body in England and national enquires, both of which produce recommendations for local organisations.[360]

Investigations should have explicit explanation of their purpose from the outset. There may be multiple requirements of the investigation, such as identifying opportunities to reduce risk and improve safety and provide patients and or families with an explanation of the events leading to the incident.[300,303,329]

Clarity is needed around language and terminology. This would benefit those involved in investigations, reading reports or evaluating the process. Investigations and hospitals should provide definitions of key terms such as recommendations and ensure consistency in their use.

Investigators should be encouraged to "show their working" within investigations.[369:p.366] The scoping review demonstrated a lack of details reported about the methodological approaches and considerations within hospital investigations. Openly reporting methods and a descriptive account of the investigation approach will allow evaluation, critique and ultimately the potential for iterative improvement within the peer-reviewed literature and public domain.[369]

An awareness of the different and sometimes conflicting views of the public, staff and experts, highlighted in the scenario study, is important. This will prepare those conducting and overseeing investigations for the discussions they may need to have with these groups during an investigation. Investigators will need to be skilled in managing these dynamics.

While the bias research in the context of healthcare investigations is relatively underdeveloped, there is a growing evidence base of approaches that may mitigate potential negative impacts of bias. This thesis demonstrates the potentially significant impact of cognitive bias on the investigation process. Hospitals and investigators should explicitly consider the impacts of bias and seek out evidenced-based approaches for mitigation. It may be prudent for investigations to include statements concerning bias within reports.

Investigators and those with oversight for investigation processes should receive evidence-based training. A consideration of the impact of bias could be built into checks and mechanisms in a similar way that tools are being used to evaluate health equity consideration within maternity and newborn safety investigations.[370] Training by itself is unlikely to mitigate the impacts of bias but may be effective when combined with other strategies like that, and others discussed briefly below.[272] It is therefore important that hospitals are aware that making investigators aware of bias will not prevent their susceptibility to it and potential impact on investigations.

The Delphi criteria could be used to reflect on the current approach to investigation and recommendation analysis, potentially identifying opportunities for improvement in the process. As Leistikow and colleagues argued, evaluation of the learning process may be as beneficial as the measuring outcomes.[211] In other words a focus on the components of the investigation process, as presented in the Delphi study, may provide organisations with a manageable approach to evaluating and improving practice. The criteria within the Delphi, drawn from a large body of supporting literature and empirically developed with experts provides a valid standard against which to examine our current investigation practice.

6.4 Implications for policy

National and local policy documents set expectations and standards for investigations, and can therefore have a significant impact on practice.[371,372] As a highly relevant example a recent document analysis of incident investigation policy documents in England found a surprising lack of guidance or detail relating to the involvement of patients and families in investigations.[333] Addressing these deficiencies in policy may contribute to improved organisational learning and reduced compound harm for those involved in investigations.[48,333] This PhD and thesis presents a number of implications to consider in policy decisions as well as the related to documents and guidance.

Policy documents should be clear for those using them. Significant confusion and lack of agreement around terminology was highlighted in the scoping review. Recent documents in the literature continue to use a range of different terms with overlapping meaning. As an example, PSIRF described the identification of areas for improvement and safety actions,[161] the London Protocol uses the term recommendations and others the term risk controls.[98,255] While it is encouraging to see that clear definitions of these terms may be provided, these differences are concerning for disagreements or confusion about the nature and purpose of recommendations. Work could be undertaken to achieve consensus in this area and provide a more universal language in relation to investigations and recommendations. There may be benefit from providing standards or expectations in relation to the reporting of investigations such as the format and content of reports. In the same way that research papers have expected structure and content to ensure that the purpose, approach and analysis and transparent and repeatable, so too could investigation reports benefit from this kind of reporting.

The importance of the selection of investigators should be discussed within policy. The scenario study has significant implications for the conduct of investigations and selection of investigators. In reality, it is likely that individual investigators are limited in their ability to mitigate bias even when provided with specific training and therefore systemic factors, potentially beyond their control, will need to be considered.[272] Organisational factors or strategies might include what information is provide to investigators and when and arrangements for independent verification.[272]

Policy should, where possible, use the best available evidence. Those deciding on policy in this area might benefit from engaging with and supporting current research to integrate evidence-based approaches to bias mitigation. The Learn Together project is a good example of this kind of rapid integration between research and policy, that could be emulated within other areas of the investigation process, such as bias.[328]

Policy needs to explicitly attend to the human tendency to want to investigate when harm is severe. While patient harm plays an important part in individuals' responses to an incident, policymakers will need to consider when harm severity is justification for investigation and when it is not. This is of particular relevance to current policy in England which encourages proportionate responses to incidents. The scenario study demonstrates that not only do people tend to want to investigate rather than not investigate and incident, but also that this tendency is much greater as harm increases.

The competence and capacity of those investigating should be discussed in policy. Investigations should be led by those with expertise and experience and would benefit from defined competencies for investigators. This is a topic that has received increasing attention over the last few years, with calls for the professionalisation of investigators.(refs)The delphi criteria could be used to inform standards and the content of policy documents at a national, regional and local hospital level. As an example, the Dutch national incident reporting system utilizes an internally developed set of criteria to evaluate investigation learning processes within hospitals.[211] Leistikow and colleagues raise an important question, which would also apply to the criteria within the delphi in this PhD, "whether the quality of SE analysis reports is a true reflection of a hospital's learning process".[211:p.255]

6.5 Implications for Research

Academics should use consistent and unified terms in the literature. The suggestions around clarity and consistency of terminology in relation to investigations and recommendations, discussed in the previous sections, seem as relevant to the research domain. Moving towards a universal language in this context would make research more generalisable and globally relevant.

Rigorous empirical evaluation and healthcare adaptation of investigation approaches is needed. Incident investigations remain foundational to patient safety measurement and improvement but more enquiry is needed about their

effectiveness or impact. As alternatives to traditional formal investigations, such as after-action reviews, are increasingly encouraged research will need to consider the effectiveness of these. New approaches and tools for recommendation generation are more likely to be successful if adapted and designed relative to the unique and complex context of health care, with rigorous empirical evaluation. Further specific research is needed to continue to empirically develop frameworks of contributory factors that themselves impact the findings and recommendations generated by investigations.[144]

Appropriate measures of safety and investigation impact need to be defined. More research is needed to consider specifically what measures are appropriate for measuring recommendation or investigation effectiveness.

The potential negative impact of recommendation implementation needs to be evaluated. The scenario study highlighted the tendency of investigators to select recommendations rather than not. With the potential of safety investigations to contribute to the creation of safety clutter or low-value safety practices,[277,278] further research is needed to explore the potential negative impact of investigations and recommendations in this context.

The impacts of bias and potential mitigation strategies should be a research priority. Research is needed to develop and empirically test strategies to mitigate the impact of cognitive biases on investigations, such as investigator training, blinding and unmasking (e.g. to patient outcome) and independent verification.[272] Future research is needed to explore the impact of bias on alternatives to traditional investigation in patient safety, such as After-Action Review, 'SWARM' and Structured Judgement Review.

Further research is needed to make the criteria presented in chapter 5 usable in the real world. The Delphi criteria represent an empirical step forward in defining quality in investigations and recommendations but further research is needed to 1) identify gaps such as criteria relating to inequity,[351,252] 2) test and develop their usability in real world settings, and 3) work with stakeholders to develop explanatory notes. Future research might also

examine, using the delphi criteria, the relationship between investigation quality, recommendation quality, and outcome measures relating to safety.

The importance of national and local policy documents in translating best evidence approaches to investigations,[328,333] a documentary analysis of hospital policy documents in relation to investigation and recommendation quality may be of benefit. This would allow a gap analysis and highlight opportunities to improve local policy documents.

6.6 Strengths and limitations of the thesis

Chapters three, four and five, each contain discussion of the strengths and limitations of the individual studies. This section will consider several strengths and limitations of the thesis as a whole in addressing the aims of the PhD.

6.6.1 Generalisability and relevance

One key strength of this thesis is the potential generalisability of the study findings. As illustrated in the scoping review the approaches to investigation, such as RCA, are similar across other health systems in the USA, Australia, Netherlands, Brazile, Hong Kong. In fact the problems and challenges faced by those conducting investigations are similar across countries,[254] therefore the findings and analysis within this thesis may have international relevance.

6.6.2 The black box of recommendation generation remains

The scoping review presented important findings relating to the types of recommendations proposed in hospitals as well as consideration about their quality, but did not reveal how hospitals actually generate them, which was one of the objectives of this PhD. The rationale for the choice of scoping review as a method to answer this question was provided in section 2.5.1. As the scoping review was unable to provide an answer to this question it is important to explain why further studies where not undertaken to answer this question within

the PhD. This gap in the literature is not peculiar to healthcare. In a study of Swedish investigators (including those from healthcare, nuclear, and transportation) only 5% of respondents referred to explicit methods or strategies for generating recommendations.[373] The remainder of investigators either responded 'no' to the question "*Do you use any particular method or strategy in the process of formulating the recommendations?"*, or "were vague and referred to general criteria such as; 'realistic recommendations,' or that the recommendations are 'derived from the finding[s]'.".[373] To the PhD candidates knowledge, there remains no studies in the literature specifically examining how hospitals generate recommendations.

6.7 Reflection on professional identity

Neary described professional identity as fluid rather than static, and influenced not only by how we see ourselves, but also how we think others perceive us, and the perceptions of society.[374,375] While there is variation in the definitions of professional identity in the literature, Fitzgerald suggests the following characteristics:

> *"the ability to perform the functions of the profession; knowledge, as evidenced by education and/or certification; identification with a community of practice and with the values and ethics of the profession; and personal identification as a professional within an identified professional group."*[376]

My professional identify has changed during the course of my PhD, and continues to vary depending on where I am (clinical or academic environments) and who I am talking to. If working clinically, and talking to patients or families, I continue to introduce myself as a doctor and do not discuss my academic activity. My research in patient safety is actually highly relevant to patients, families, and staff in the hospital I work in, but I find that 'patient safety' is not routinely discussed in day-to-day clinical work, unless there has been a serious breach. When speaking to staff in the hospital who work in patient safety or

quality improvement, or to those within academic environments I would, earlier in my PhD, introduce myself as a doctor doing a PhD in patient safety. As I have approached the end of my PhD I have described myself as a doctor who does patient safety research. On reflection it is clear I identify as a clinical doctor first, and an academic second. While the term clinical academic is commonly used in the UK to describe someone who combines both clinical and academic work, I have only recently started to consider myself one; this may reflect a slow transition to accepting a dual professional identity, rather than two separate identities.

6.8 Conclusion

The investigation of patient safety incidents and generation of recommendations is a highly complex amalgam of social, political, and technical activities, which must examine the highly complex systems that make up healthcare delivery. This thesis has highlighted the widespread tendency of individual focused recommendations and confusion and disagreement about how recommendations should be generated or judged for effectiveness. The importance and potentially significant impact of investigator outcome bias has been demonstrated along with the important differences between how the public, staff and patient safety experts react to incidents and select recommendations. The Delphi study represents an important empirically grounded step forward in defining a good quality investigation and recommendations. Taken together the findings of this PhD can inform evidenced based practical improvements and further research to improve the effectiveness of recommendations and investigations.

Healthcare has yet to realise the potential of investigations to improve safety. If specific system-oriented methodologies were used, with a focus on a smaller number of quality investigations, conducted by appropriately trained people significant improvements might be achieved.

References

1. Department of Health. An organisation with a memory: report of an expert group on learning from adverse events in the NHS. London: The Stationery Office, 2000.
2. Kohn LT, Corrigan JM, Donaldson MS (Institute of Medicine). To err is human: building a safer health system. Washington, DC: National Academy Press, 2000.
3. Macrae C. The problem with incident reporting. *BMJ Qual Saf*. 2016;25:71–75.
4. Rex JH, Turnbull JE, Allen SJ, et al. Systematic root cause analysis of adverse drug events in a tertiary referral hospital. The Joint Commission journal on quality improvement 2000;26(10):563-575.
5. Perkins JD, Levy AE, Duncan JB, et al. Using root cause analysis to improve survival in a liver transplant program. Journal of Surgical Research 2005;129(1):6 16.
6. Bromiley M. Have you ever made a mistake? RCoA Bulletin 2008;48:2442–5.
7. National Patient Safety Agency. Haemodialysis patients: risks associated with water supply (hydrogen peroxide). NPSA, 2008. http://www.nrls.npsa.nhs.uk/resources/?EntryId45=59893 (accessed 18 March 2025)
8. Loh HP, de Korne DF, Chee SP, et al. Reducing wrong intraocular lens implants in cataract surgery: 3 years of experience with the SEIPS framework in Singapore. International journal of health care quality assurance 2017;30(6):492 505.
9. Lundberg J, Rollenhagen C, Hollnagel E. What you find is not always what you fix--how other aspects than causes of accidents decide recommendations for remedial actions. Accid Anal Prev. 2010 Nov;42(6):2132-9.

10. Card AJ, Ward J, Clarkson PJ. Successful risk assessment may not always lead to successful risk control: a systematic literature review of risk control after root cause analysis. *J Healthc Risk Manag*. 2012;31:6–12.
11. Wrigstad J, Bergström J, Gustafson P. Mind the gap between recommendation and implementation—principles and lessons in the aftermath of incident investigations: a semi-quantitative and qualitative study of factors leading to the successful implementation of recommendations. BMJ Open 2014;4:e005326.
12. Peerally MF, Carr S, Waring J, et al. The problem with root cause analysis. BMJ Quality & Safety 2017;26:417-422.
13. Braithwaite J. Changing how we think about healthcare improvement. BMJ 2018; 361 :k2014
14. Martin-Delgado J, Martínez-García A, Aranaz JM, Valencia-Martín JL, Mira JJ. How Much of Root Cause Analysis Translates into Improved Patient Safety: A Systematic Review. Med Princ Pract. 2020;29(6):524-531.
15. Lundberg J, Rollenhagen C, Hollnagel E. What-You-Look-For-Is-What-You-Find – The consequences of underlying accident models in eight accident investigation manuals. Safety Science. Volume 47, Issue 10, December 2009, Pages 1297-1311
16. HSIB. How we improve patient safety. 2022. Available at: https://www.hsib.org.uk/. Accessed October 1, 2022.
17. UKOM. About Ukom. 2021. Available at: https://ukom.no/om-ukom. Accessed October 1, 2022.
18. Kirkup B. The Report of the Morecambe Bay Investigation. 2015. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/408480/47487_MBI_Accessible_v0.1.pdf. Accessed November 1, 2022.
19. Nightingale F. Notes on hospitals. London. Parker and Son; 1863.
20. Schimmel EM. The hazards of hospitalization. Ann Intern Med. 1964;60:100-110.

21. Steel K, Gertman PM, Crescenzi C, et al. Iatrogenic illness on a general medical service at a university hospital. N Engl J Med. 1981;304:638-642.
22. Vincent CA. Research into medical accidents: a case of negligence? BMJ. 1989 Nov 4;299(6708):1150-3.
23. Pinkus RL. Mistakes as a social construct: an historical approach. Kennedy Inst Ethics J 2001;11:117-33.
24. Bedell SE, Deitz DC, Leeman D, Delbanco TL. Incidence and characteristics of preventable iatrogenic cardiac arrests. JAMA. 1991;265:2815-2820.
25. Brennan TA, Leape LL, Laird N, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. N Engl J Med. 1991;324:370-376.
26. Leape LL, Brennan TA, LairdN, et al. The nature of adverse events in hospitalized patients: Results of the Harvard Medical Practice Study II. N Engl J Med. 1991;324:377-384.
27. Leape LL. Error in Medicine. JAMA. 1994;272(23):1851–1857.
28. Blendon RJ, DesRoches CM, Brodie M, et al. Views of practicing physicians and the public on medical errors. N Engl J Med 2002;347:1933–40.
29. Stelfox HT, Palmisani S, Scurlock C*, et al*The "*To Err is Human*" report and the patient safety literature *BMJ Quality & Safety* 2006;**15:**174-178.
30. Emanuel L, Berwick D, Conway J, et al. What Exactly Is Patient Safety? In: Henriksen K, Battles JB, Keyes MA, et al., editors. Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 1: Assessment). Rockville (MD): Agency for Healthcare Research and Quality; 2008 Aug. Available from: https://www.ncbi.nlm.nih.gov/books/NBK43629/
31. Baker G, Norton P, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian adverse events study: the incidence of adverse events among hospital patients in Canada.CMAJ 2004;170:1678-86.
32. Sari AB, Sheldon TA, Cracknell A, et al. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. BMJ. 2007;334:79.

33. Hiatt HH, Barnes BA, Brennan TA, et al. A study of medical injury and medical malpractice. N Engl J Med 1989;321:480–4.
34. Wilson R, Runciman W, Gibberd R, Harrison B, Newby L, Hamilton J. The quality in Australian health care study. Med J Aust 1995;163:458-71.
35. Davis P, Lay YR, Briant R, Ali W, Scott A, Schug S. Adverse events in New Zealand public hospitals I: occurrence and impact. NZMedJ2002;115:U271.
36. Panagioti M, Khan K, Keers R N, Abuzour A, Phipps D, Kontopantelis E et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. BMJ 2019; 366:l4185.
37. Bates et al. The Safety of Inpatient Health Care. N Engl J Med. 2023 Jan 12;388(2):142-153.
38. Global patient safety report 2024. Geneva: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO.
39. Cooper J, Edwards A, Williams H, Sheikh A, Parry G, Hibbert P, Butlin A, Donaldson L, Carson-Stevens A. Nature of Blame in Patient Safety Incident Reports: Mixed Methods Analysis of a National Database. Ann Fam Med. 2017 Sep;15(5):455-461.
40. Ramsey L, Sheard L, Waring J, McHugh S, Simms-Ellis R, Louch G, Ludwin K, O'Hara JK. Humanizing processes after harm part 1: patient safety incident investigations, litigation and the experiences of those affected. Front Health Serv. 2025 Jan 3;4:1473256.
41. Massó Guijarro P, Aranaz Andrés JM, Mira JJ, et al. Adverse events in hospitals: the patient's point of view. BMJ Quality & Safety 2010;19:144-147.
42. Edrees HH, Paine LA, Feroli ER, Wu AW. Health care workers as second victims of medical errors. Pol Arch Med Wewn. 2011 Apr;121(4):101-8.
43. Burlison JD, Scott SD, Browne EK, Thompson SG, Hoffman JM. The Second Victim Experience and Support Tool: Validation of an Organizational Resource for Assessing Second Victim Effects and the Quality of Support Resources. J Patient Saf. 2017 Jun;13(2):93-102.

44. Cabilan CJ, Kynoch K. Experiences of and support for nurses as second victims of adverse nursing errors: a qualitative systematic review. JBI Database System Rev Implement Rep. 2017 Sep;15(9):2333-2364.
45. Willis D, Yarker J, Lewis R. Lessons for leadership and culture when doctors become second victims: a systematic literature review. BMJ Leader 2019;3:81-91.
46. McQueen JM, Gibson KR, Manson M, Francis M. Adverse event reviews in healthcare: what matters to patients and their family? A qualitative study exploring the perspective of patients and family. BMJ Open. 2022 May 9;12(5):e060158.
47. Hodkinson A, Zhou A, Johnson J, Geraghty K, Riley R, Zhou A, Panagopoulou et al. Associations of physician burnout with career engagement and quality of patient care: systematic review and meta-analysis. BMJ. 2022;378:e070442.
48. Ramsey L, McHugh S, Simms-Ellis R, et al. Patient and Family Involvement in Serious Incident Investigations From the Perspectives of Key Stakeholders: A Review of the Qualitative Evidence. J Patient Saf 2022;18:e1203–10
49. Ramsey L, Hughes J, Hazeldine D, Seddon S, Gould M, Wailling J, Murray J, McHugh S, Simms-Ellis R, Halligan D, Ludwin K, O'Hara JK. Humanising processes after harm part 2: compounded harm experienced by patients and their families after safety incidents. Front Health Serv. 2024 Dec 17;4:1473296.
50. Slawomirski L, Auraaen A, Klazinga N. The Economics of patient safety: Strengthening a value-based approach to reducing patient harm at national level. OECD Health Working Papers, No. 96. Paris: Organisation for Economic Co-operation and Development; 2017
51. Slawomirski L, Klazinga N. The economics of patient safety: from analysis to action. OECD Health Working Papers No. 145. Paris: Organisation for Economic Co-operation and Development; 2022

52. Jha AK, Larizgoitia I, Audera-Lopez C, Prasopa-Plaizier N, Waters H, Bates DW. The global burden of unsafe medical care: analytic modelling of observational studies. BMJ Qual Saf. 2013;22:809–15.
53. Chauhan A, Walton M, Manias E, Walpola RL, Seale H, Latanik M et al. The safety of health care for ethnic minority patients: a systematic review. Int J Equity Health. 2020;19:118. doi:10.1186/s12939-020-01223-2
54. Gangopadhyaya A. Black patients are more likely than white patients to be in hospitals with worse patient safety conditions. Washington, DC: Urban Institute; 2021 (https://www.rwjf.org/content/dam/farm/reports/issue_briefs/2021/rwjf464731, accessed 28-1-25)
55. Wade C, Malhotra A M, McGuire P, Vincent C, Fowler A. Action on patient safety can reduce health inequalities *BMJ* 2022; 376 :e067090
56. Strengthening primary health care to tackle racial discrimination, promote intercultural services and reduce health inequities. Geneva: World Health Organization; 2022 (https://iris.who.int/handle/10665/363854, accessed 28-1-25)
57. Knight M, Bunch K, Felker A, Patel R, Kotnis R, Kenyon S, Kurinczuk JJ (Eds.) on behalf of MBRRACE-UK. Saving Lives, Improving Mothers' Care Core Report - Lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2019-21. Oxford: National Perinatal Epidemiology Unit, University of Oxford 2023.
58. Doupi P. National reporting systems for patient safety incidents: a review of the situation in Europe. Helsinki: National Institute for Health and Welfare, 2009.
59. Runciman W B. Iatrogenic injury in Australia. A report by the Australian Patient Safety Foundation for the National Health Priorities and Quality Branch. Canberra, 2000. Available at: http://www.apsf.net.au/dbfiles/Iatrogenic_Injury.pdf (accessed 31 Mar 2007)

60. Barach P, Small S D. Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. BMJ 2000 320759–763.
61. Dunn D. Incident reports—their purpose and scope. AORN J. 2003;78: 46, 49–61, 65–6.
62. Shaw R, Drever F, Hughes H.et al Adverse events and near miss reporting in the NHS. Qual Saf Health Care 2005 14279–283.
63. Nuckols TK, Bell DS, Liu H, et al. Rates and types of events reported to established incident reporting systems in two US hospitals. Qual Saf Health Care. 2007;16:164–168.
64. Fitzgerald EM, Cawley D, Rowan NJ. Irish staff nurses perceptions of clinical incident reporting. Int J Nurs Di dact. 2013;3(2):14-21.
65. The Joint Commission (2025) Comprehensive Accreditation Manual for Hospitals: Sentinel Event Policy. Available at https://www.jointcommission.org/-/media/tjc/documents/resources/patient-safety-topics/sentinel-event/2025/se-3-camhospital_se_jan_2025.pdf (last accessed 10/05/2025)
66. Carter AW, Mossialos E, Darzi A (2015) A national incident reporting and learning system in England and Wales, but at what cost?, Expert Review of Pharmacoeconomics & Outcomes Research, 15:3, 365-368
67. Mushcab H, Bunting D, Yami S, Abandi A, Hunt C. An evaluation of Datix implementation for incident reporting at Johns Hopkins Aramco Healthcare. Journal of Patient Safety and Risk Management. 2020;25(2):67-74.
68. Owusu R, Sathiavageeswaran M. Datix reporting in University Hospitals of North Midlands: what can trainees learn from this? Future Healthc J. 2023 Nov;10(Suppl 3):118-119.
69. NHS England. Serious incident framework. Supporting Learning to Prevent Recurrence. 2015. http://www.england. nhs.uk/wp-content/uploads/2015/04/serious-incidnt-framwrk upd.pdf (accessed 31 Mar 2025).

70. Australian Commission on Safety and Quality in Health Care. Incident Management Guide. Sydney: ACSQHC; 2021.
71. Archer S, Hull L, Soukup T, et al. Development of a theoretical framework of factors affecting patient safety incident reporting: a theoretical review of the literature. BMJ Open 2017;7:e017155.
72. O'Neil AC, Petersen LA, Cook EF, Bates DW, Lee TH, Brennan TA. Physician reporting compared with medical-record review to identify adverse medical events. Ann Intern Med. 1993 Sep 1;119(5):370-6.
73. Stanhope N, Crowley-Murphy M, Vincent C, et al. An evaluation of adverse incident reporting. J Eval Clin Pract. 1999;5:5–12.
74. Wanzel KR, Jamieson CG, Bohnen JM. Complications on a general surgery service: incidence and reporting. Can J Surg. 2000 Apr;43(2):113-7.
75. Vincent C. Incident reporting and patient safety. BMJ 2007;334:51.
76. Olsen S, Neale G, Schwab K, Psaila B, Patel T, Chapman EJ, Vincent C. Hospital staff should use more than one method to detect adverse events and potential adverse events: incident reporting, pharmacist surveillance and local real-time record review may all have a place. Qual Saf Health Care. 2007 Feb;16(1):40-4.
77. Hutchinson A, Young TA, Cooper KL, et al. Trends in healthcare incident reporting and relationship to safety and quality data in acute hospitals: results from the National Reporting and Learning System. BMJ Quality & Safety 2009;18:5-10.
78. Yu A, Flott K, Chainani N, Fontana G, Darzi A. Patient safety 2030. 2016. www.imperial.ac.uk/media/imperial-college/institute-of-global-health-innovation/centre-for-health-policy/Patient-Safety-2030-Report-VFinal.pdf.
79. Hibbert PD, Molloy CJ, Schultz TJ, Carson-Stevens A, Braithwaite J. Comparing rates of adverse events detected in incident reporting and the Global Trigger Tool: a systematic review. Int J Qual Health Care. 2023 Jul 25;35(3):mzad056.

80. NHS NRLS 2022. NRLS national patient safety incident reports: commentary October 2022https://www.england.nhs.uk/wp-content/uploads/2022/10/NAPSIR-commentary-Oct-22-FINAL-v4.pdf accessed 9-3-25
81. Sinclair A, Guérin A, Robin C, Dey P. Investigating the cost and efficiency of incident reporting in a specialist paediatric NHS hospital and impact on patient safety. Eur J Hosp Pharm. 2017 Mar;24(2):91-95.
82. Hagley G, Mills PD, Watts BV, Wu AW. Review of alternatives to root cause analysis: developing a robust system for incident report analysis. BMJ Open Qual. 2019 Aug 1;8(3):e000646.
83. Blanchfield BB, Acharya B, Mort E. The Hidden Cost of Regulation: The Administrative Cost of Reporting Serious Reportable Events. Jt Comm J Qual Patient Saf. 2018 Apr;44(4):212-218.
84. Shojania KG. The frustrating case of incident-reporting systems. BMJ Quality & Safety 2008;17:400-402.
85. Noble DJ, Pronovost PJ. Underreporting of patient safety incidents reduces health care's ability to quantify and accurately measure harm reduction. J Patient Saf 2010;6:247–50.
86. Lawton R, Parker D. Barriers to incident reporting in a healthcare system. BMJ Quality & Safety 2002;11:15-18.
87. Reason J. Human error: models and management. BMJ. 2000 Mar 18;320(7237): 768-70.
88. Mitchell I, Schuster A, Smith K, et al. Patient safety incident reporting: a qualitative study of thoughts and perceptions of experts 15 years after ‘To Err is Human’. BMJ Quality & Safety 2016;25:92-99.
89. Maxton F, Darbyshire P, Daniel DJM, Walvin T. Nursing can help end the travesty of 'Datix abuse'. J Clin Nurs. 2021 May;30(9-10):e41-e44.
90. Macrae C. Close calls: managing risk and resilience in airline flight safety. London: Palgrave, 2014.
91. Nakajima K, Kurata Y, Takeda HA web-based incident reporting system and multidisciplinary collaborative projects for patient safety in a Japanese hospital. BMJ Quality & Safety 2005;14:123-129.

92. Kumah A, Zon J, Obot E, Yaw TK, Nketsiah E, Bobie SA. Using Incident Reporting Systems to Improve Patient Safety and Quality of Care. Glob J Qual Saf Healthc. 2024 Feb 28;7(4):228-231.
93. Irwin A, Ross J, Seaton J, Mearns K (2011) Retrospective analysis of DATIX dispensing error reports from Scottish NHS Hospitals. International Journal of Pharmacy Practice, 19, pp.417-423
94. Zeng J, Nyflot MJ, Jordan LE, Sponseller PA, Novak A, Carlson J, Ermoian RP, Kane GM, Ford EC (2016) Best practices for safety improvement through high volume institutional incident learning: lessons learned from 2 years. J Radiat Oncol (2016) 5:323-333
95. Corwin GS, Mills PD, Shanawani H, Hemphill RR (2017) Root Cause Analysis of ICU Adverse Events in the Veterans Health Administration. The Joint Commission Journal on Quality and Patient Safety 2017; 43:580-590
96. Figueiredo ML, Oliveira e Silva CS, Brito MFSF, D'Innocenzo M. Analysis of incidents notified in a general hospital. Rev Bras Enferm [Internet]. 2018;71(1):111-9.
97. Hamilton MJ, McEniery JA, Osborne JM, Coulthard MG (2019) Implementation and strength of root cause analysis recommendations following serious adverse events involving paediatric patients in the Queensland public health system between 2012 and 2014. Journal of Paediatrics and Child Health 55 (2019) 1070-1076
98. Vincent C, Adams S, Bellandi T, Higham H, Michel P, Staines A. Systems Analysis of Clinical Incidents: The London Protocol 2024.
99. Kellogg KM, Hettinger Z, Shah M, Wears RL, Sellers CR, Squires M, Fairbanks RJ. Our current approach to root cause analysis: is it contributing to our failure to improve patient safety? BMJ Qual Saf. 2017 May;26(5):381-387.
100. Hibbert PD, Thomas MJW, Deakin A, et al. Are root cause analyses recommendations effective and sustainable? An observational study. *Int J Qual Health Care*. 2018;30:124–131.

101. Kwok YTA, Mah AP, Pang KM. Our first review: an evaluation of effectiveness of root cause analysis recommendations in Hong Kong public hospitals. *BMC Health Serv Res*. 2020;20:507.

102. NHS England. Patient safety data [internet]. Last accessed 29th July 2025. Available from: https://www.england.nhs.uk/statistics/statistical-work-areas/patient-safety-data/

103. NHS England. Patient safety incident response framework and supporting guidance [internet] 2024. Last accessed 29th July 2025. Available from: https://www.england.nhs.uk/publication/patient-safety-incident-response-framework-and-supporting-guidance/

104. NHS Improving Quality. Learning Handbook: After Action Review. Last accessed 29th July 2025. Available from: https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2015/08/learning-handbook-after-action-review.pdf

105. NHS England. Patient Safety Incident Response Framework supporting guidance: Guide to responding proportionately to patient safety incidents. 2022. Last accessed 29th July 2025. Available from: https://www.england.nhs.uk/wp-content/uploads/2022/08/b1465-3.-guide-to-responding-proportionately-to-patient-safety-incidents-v1.2.pdf

106. NHS Improvement. Never Events policy and framework. 2018. Last accessed 29th July 2025. Available from https://www.england.nhs.uk/wp-content/uploads/2020/11/Revised-Never-Events-policy-and-framework-FINAL.pdf

107. Woloshynowych M, Rogers S, Taylor-Adams S, et al. The investigation and analysis of critical incidents and adverse events in healthcare. Health Technol Assess. 2005;9:1–143, iii.

108. Carayon P, Bass E, Bellandi T, Gurses A, Hallbeck S, Mollo V. Socio-Technical Systems Analysis in Health Care: A Research Agenda. IIE Trans Healthc Syst Eng. 2011;1(1):145-160. doi:

10.1080/19488300.2011.619158. Epub 2011 Dec 2. PMID: 22611480; PMCID: PMC3351758.

109. Effken JA. Different lenses, improved outcomes: A new approach to the analysis and design of healthcare information systems. International Journal of Medical Informatics. 2002; 65(1):59–74.

110. Shorrock ST, Williams C. Human factors and ergonomics in practice: Improving system performance and human well-being in the real world. Boca Raton, FL: CRC Press, 2017, p.422.

111. Hollnagel, E. (2004) Barriers and Accident Prevention: Aldershot: Ashgate.

112. Toft Y, Dell G, Klockner K, Hutton A. (2012) Models of causation: Safety. In HaSPA (Health and Safety Professionals Alliance), The Core Body of Knowledge for Generalist OHS Professionals. Tullamarine, VIC. Safety Institute of Australia.

113. Heinrich, H. W. (1931). Industrial Accident Prevention: A scientific approach. New York: McGraw-Hill.

114. Bird, F. E. J., &Germain, G. L. (1985). Practical Loss Control Leadership. Loganville, Georgia: International Loss Control Institute, Inc.

115. Reason JT. Managing the risks of organizational accidents. : Ashgate Aldershot; 1997.

116. Reason J. Human Error, New York, Cambridge University Press, 1990.

117. Reason J, Parker D, Free R. Bending the Rules: The varieties, origins and management of safety violations, (Leiden: Faculty of Social Sciences, University of Leiden, 1994)

118. Lawton, R. (1998). Not working to rule: understanding procedural violations at work. *Safety science*, *28*(2), 77-95.

119. Vincent C, Taylor-Adams S, Stanhope N. Framework for analysing risk and safety in clinical medicine *BMJ* 1998; 316 :1154

120. Reason J. Understanding adverse events: human factors. BMJ Quality & Safety 1995;4:80-89.

121. Runciman WB, Sellen A, Webb RK, et al. The Australian incident monitoring study. Errors, incidents and accidents in anaesthetic practice. Anaesth Intensive Care 1993;21:506–9.

122. Van Vuuren W, Shea C, van der Schaaf TW. The Development of an Incident Analysis Tool for the Medical Field. Eindhoven: Report from Faculty of Technology Management, University of Eindhoven, 1997.

123. Bagian JP, Gosbee J, Lee CZ, et al. The Veterans Affairs root cause analysis system in action. Jt Comm J Qual Improv 2002;28:531–45.

124. Taylor-Adams S, Vincent C. Systems analysis of clinical incidents: the London Protocol. Clin Risk 2004;10:211–20.

125. The conceptual framework for the international classification for patient safety. Geneva: World Health Organization; 2009 (https://www.who.int/ publications/i/item/the-conceptual-framework-for the-international-classification-for-patient-safety (icps), accessed 12 July 2021)

126. Benner, L. (1975). Accident Investigations: Multilinear Events Sequencing Methods. Journal of Safety Research, 7(2), 67-73.

127. Benner L. Accident Models: How Underlying Differences Affect Workplace Safety. Journal of Occupational Accidents, 6 (1984) 127-131

128. Rasmussen, J. (1990). The role of error in organizing behaviour. *Ergonomics, 33*(10-11), 1185–1199.

129. Leveson, N. (2004). A new accident model for engineering safer systems. Safety Science, 42, 237-270.

130. Perrow C. Normal accident theory, living with high risk technology. : New York: Basic Books; 1984.

131. Tenner, E. (1996). Why things bite back. London: Fourth Estate Limited.

132. Hollnagel, E. (1993). Human reliability analysis: Context and control. London: Academic Press.

133. Hollnagel, E. (1998). Cognitive reliability and error analysis method: CREAM: Elsevier.

134. Hollnagel, E., Goteman, Ö., 2004. The Functional Resonance Accident Model. Proc. 2014 Cogn. Syst. Eng. Process plant 155–161.

135. Hollnagel, E. (2005). The ETTO Principle - Efficiency-Thoroughness Trade Off. Retrieved from http://www.ida.liu.se/~eriho/ETTO_M.htm

136. Hulme, A., Stanton, N. A., Walker, G. H., Waterson, P., & Salmon, P. M. (2021). Are accident analysis methods fit for purpose? Testing the criterion-referenced concurrent validity of AcciMap, STAMP-CAST and AcciNet. Safety Science, 144, Article 105454.

137. Hollnagel E, Braithwaite J, Wears RL. *Resilient health care*. Surrey England: Ashgate Publishing Limited, 2013.

138. Woodward S. Moving towards a safety II approach. Journal of Patient Safety and Risk Management. 2019;24(3):96-99.

139. Verhagen MJ, de Vos MS, Sujan M, et al. The problem with making Safety-II work in healthcare. BMJ Quality & Safety 2022;31:402-408.

140. Reason, JT. Understanding adverse events: the human factor. In Vincent, CA. Clinical Risk Management. Enhancing patient safety (2nd edition). London, BMJ Books, 2001; pp9 30.

141. Diller T, Helmrich G, Dunning S, et al. The Human Factors Analysis and Classification System (HFACS) applied to health care. American Journal of Medical Quality 2014;29:181–90.

142. Gaba DM. Structural and organizational issues in patient safety: a comparison of health care to other high-hazard industries. Calif Manage Rev 2000;43:83–102.

143. Vincent C. Patient Safety. Oxford: BMJ Books, 2010

144. Lawton R, McEachan RRC, Giles SJ, et al. Development of an evidence-based framework of factors contributing to patient safety incidents in hospital settings: a systematic review. BMJ Quality & Safety 2012;21:369-380.

145. Peerally MF, Carr S, Waring J, Martin G, Dixon-Woods M. A content analysis of contributory factors reported in serious incident investigation reports in hospital care. Clin Med (Lond). 2022 Sep;22(5):423-433.

146. Vincent CA Analysis of clinical incidents: a window on the system not a search for root causes BMJ Quality & Safety 2004;13:242-243.

147. Braithwaite J, Westbrook MT, Mallock NA, et al. Experiences of health professionals who conducted root cause analyses after undergoing a safety improvement programme. Qual Saf Health Care 2006;15:393–9.

148. Taitz J, Genn K, Brooks V, et al. System-wide learning from root cause analysis: a report from the New South Wales Root Cause Analysis Review Committee. Quality & safety in health care 2010 Dec;19(6):e63.

149. Duchscherer C, Davies J. Systematic Systems Analysis: A practical Approach to Patient Safety Reviews. 2012

150. Nygren M, Roback K, Öhrn A, et al. Factors influencing patient safety in Sweden: perceptions of patient safety officers in the county councils. BMC health services research 2013;13(1):52.

151. Croteau RJ. Root Cause Analysis in Healthcare: Tools and Techniques. Oakbrook Terrace, IL : Joint Commission Resources ; 2010 .

152. Canham A, Thomas Jun G, Waterson P, et al. Integrating systemic accident analysis into patient safety incident investigation practices. Appl Ergon 2018;72:1–9.

153. Isherwood P, Waterson P. To err is system; a comparison of methodologies for the investigation of adverse outcomes in healthcare. Journal of Patient Safety and Risk Management. 2021;26(2):64-73.

154. Bigio C, Rodrigues M, de Melo C, et al. Implementation of Improvements Based on the Analysis of Severe Adverse Events in Paediatric Patients. Patient Saf 2022;4:34–42.

155. Woodier N, Whiting K, Bennett O. Investigations by acute-hospital staff: AcciMaps or HFACS. Contemp Ergon Hum Factors 2022;4:203–14.

156. Wu AW, Lipshutz AK, Pronovost PJ. Effectiveness and efficiency of root cause analysis in medicine. JAMA 2008;299:685–7.

157. Latino RJ. How is the effectiveness of root cause analysis measured in healthcare? *J Healthc Risk Manag*. 2015;35:21–30.

158. Care Quality Commission (2016) Learning from serious incidents in NHS acute hospitals: a review of the quality of investigation reports, online: http://www.cqc.org.uk/publications/themed-work/briefing-learning-seriousincidents-nhs-acute-hospitals: Care Quality Commission.

159. Nicolini D, Waring J, Mengis J. The challenges of undertaking root cause analysis in health care: a qualitative study. J Health Serv Res Policy 2011;16(Suppl 1):34–41.

160. White Paper Working Group. LEARNING FROM ADVERSE EVENT. Chartered Institute of Ergonomics & Human Factors. 2020. Available from https://ergonomics.org.uk/resource/learning-from-adverse-events.html

161. NHS England. Safety action development guide [internet]. 2022. Last accessed 29th July 2025. Available from: https://www.england.nhs.uk/wp-content/uploads/2022/08/B1465-Safety-action-development-v1.1.pdf

162. Arber, A. (2006). Reflexivity: A challenge for the researcher as practitioner? Journal of Research in Nursing, 11(2), 147–157.

163. Huberman, A. M., & Miles, M. B. (2002). The qualitative researcher's companion. SAGE Publications.

164. Milligan, L. (2014). Insider-outsider-inbetweener? Researcher positioning, participative methods and cross cultural educational research. Compare: A Journal of Comparative and International Education, 4, 1–16.

165. Chhabra, G. (2020). Insider, outsider or an in-betweener? Epistemological reflections of a legally blind researcher on conducting cross-national disability research. Scandinavian Journal of Disability Research, 22, 307–317.

166. Edelheim, J. (2014). Ontological, epistemological and axiological issues. In D. Dredge, D. Airey, & M. J. Gross (Eds.), The Routledge handbook of tourism and hospitality education (pp. 30-42). Routledge.

https://www.routledgehandbooks.com/doi/10.4324/9780203763308.ch3

167. Killam, L. (2013). Research terminology simplified: Paradigms, axiology, ontology, epistemology and methodology. https://books.google.com.kh/books/about/Research_terminology_simplified.html?id=nKMnAgAAQBAJ&redir_esc=y

168. Potter, W. J. (2017). Epistemology. In J. Matthes, S. C. Davis, & R. F. Potter (Eds.), The international encyclopedia of communication research methods (pp. 1-14). John Wiley & Sons.

169. Sol, K., and Heng, K. (2022) Understanding epistemology and its key approaches in research. Cambodian Journal of Educational Research (2022)

170. Gillespie A, Glăveanu V, de Saint Laurent C. Epistemology: How We Know. In: *Pragmatism and Methodology: Doing Research That Matters with Mixed Methods*. Cambridge University Press; 2024:26-48.

171. Biesta GJJ, Burbules NC. Pragmatism and Educational Research. Rowman & Littlefield Publishers, Inc.; 2003.

172. Feilzer YM. Doing mixed methods research pragmatically: implications for the rediscovery of pragmatism as a research paradigm. J Mix Methods Res. 2010;4(1):6-16.

173. Kelly, L. M., & Cordeiro, M. (2020). Three principles of pragmatism for research on organizational processes. *Methodological Innovations*, *13*(2).

174. Tamuz M, Franchois K, Thomas E. What's past is prologue: organizational learning from a serious patient injury. Saf Sci. 2011;49:75–82.

175. Nicolini D, Waring J, Mengis J. Policy and practice in the use of root cause analysis to investigate clinical adverse events: mind the gap. Soc Sci Med. 2011;73:217–225.

176. Hildebrand DL. Pragmatic democracy: inquiry, objectivity, and experience. Metaphilosophy. 2011;42(5):589-604.

177. Allemang B, Sitter K, Dimitropoulos G. Pragmatism as a paradigm for patient-oriented research. Health Expect. 2022 Feb;25(1):38-47.

178. Dückers M, Faber M, Cruijsberg J, et al. Safety and risk management interventions in hospitals: a systematic review of the literature. *Med Care Res Rev*. 2009;66(6 suppl):90S–119S.

179. Pham JC, Kim GR, Natterman JP, et al. ReCASTing the RCA: an improved model for performing root cause analyses. Am J Med Qual 2010;25:186–91.

180. Card AJ, Ward JR, Clarkson PJ. Generating Options for Active Risk Control (GO-ARC): introducing a novel technique. J Healthc Qual 2014;36:32–41.

181. Vincent C, Amalberti R. Strategies for safety. In: Vincent C, Amalberti R, eds. Safer Healthcare: Strategies for the Real World. Cham, Heidelberg, New York, Dordrecht, London: Springer; 2016:59–71.

182. Morse RB, Pollack MM. Root cause analyses performed in a children's hospital: events, action plan strength, and implementation rates. J Healthc Qual. 2012;34:55–61.

183. National Patient Safety Foundation, (NPSF). *RCA² Improving Root Cause Analyses and Actions to Prevent Harm*. Boston, MA: National Patient Safety Foundation; 2015.

184. CDC NIOSH. NIOSH Topic: Engineering Controls. 2010. Available at: https://www.cdc.gov/niosh/topics/hierarchy/default.html. Accessed September 1, 2021.

185. Occupational Safety and Health Act (OSH Act) of 1970. United States of America. Available online at https://www.osha.gov/laws-regs/oshact/toc. Last accessed 5-4-25

186. Hughes D. Root cause analysis: bridging the gap between ideas and execution. Top Patient Saf. 2006. Available at: https://www.ihi.org/resources/Pages/Tools/RCA2-Improving-Root-Cause-Analyses-and-Actions-to-Prevent-Harm.aspx?PostAuthRed=/resources/_layouts/download.aspx?SourceURL=/resources/Knowledge%20Center%20Assets/Tools%20-

%20RCA2ImprovingRootCauseAnalysesandActionstoPreventHarm_f65ed89f-eb0c-490f-b05a-a772a68fd6d9/RCA2_ImprovingRootCauseAnalysesandActionstoPreventHarm.pdf. Accessed September 16, 2021.

187. Kingston M, Evans S, Smith B, et al. Attitudes of doctors and nurses towards incident reporting: a qualitative analysis. Med J Aust 2004;181:36–9.

188. Evans SM, Berry JG, Smith BJ, et al. Attitudes and barriers to incident reporting: a collaborative hospital study. Qual Saf Health Care 2006;15:39–43.

189. Wachter RM, Pronovost PJ. Balancing “no blame” with accountability in patient safety. N Engl J Med 2009;361:1401–6.

190. Walster E. Assignment of responsibility for an accident. J Pers Soc Psychol 1966;3:73–9. doi:10.1037/h002273

191. Robbennolt JK. Outcome Severity and Judgements of 'Responsibility': A Meta-Analytic Review. J Appl Soc Psychol 2000;30:2575–609.

192. Caplan RA, Posner KL, Cheney FW. Effect of outcome on physician judgments of appropriateness of care. JAMA 1991;265:1957–60.

193. Meurier CE, Vincent CA, Parmar DG. Nurses’ responses to severity dependent errors: a study of the causal attributions made by nurses following an error. J Adv Nurs 1998;27:349–54.

194. Parker D, Lawton R. Judging the use of clinical protocols by fellow professionals. Soc Sci Med 2000;51:669–77.

195. Lawton R, Gardner P, Plachcinski R. Using vignettes to explore judgements of patients about safety and quality of care: the role of outcome and relationship with the care provider. Health Expect 2011;14:296–306.

196. Baron J, Hershey JC. Outcome bias in decision evaluation. J Pers Soc Psychol 1988;54:569–79.

197. Pezzo MV. Hindsight Bias: A Primer for Motivational Researchers. Social & Personality Psych 2011;5:665–78.

198. Dror IE. Cognitive and Human Factors in Expert Decision Making: Six Fallacies and the Eight Sources of Bias. Anal Chem 2020;92:7998–8004.

199. Parker D, Lawton R. Psychological contribution to the understanding of adverse events in health care. Qual Saf Health Care 2003;12:453–7.

200. Munn, Z., Peters, M.D.J., Stern, C. *et al*. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* **18**, 143 (2018).

201. Pearson A. Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews. JBI Reports. 2004;2:45–64.

202. Pearson A, Wiechula R, Court A, Lockwood C. The JBI model of evidence-based healthcare. International Journal of Evidence-Based Healthcare. 2005;3(8):207–15.

203. Herskovits M.J. (1950) The hypothetical situation: a technique ofÆeld research. Southwestern Journal of Anthropology 6, 32-40.

204. Anderson H.H. & Anderson G.L. (1951) An Introduction to Projective Techniques and Other Devices for Understanding Human Behavior. Prentice Hall, Englewood Cliffs, NJ, USA.

205. Ouslander J.G., Tymchuk A.J. & Krynski M.D. (1993) Decisions about enteral tube feeding amongst the elderly. Journal of the American Geriatrics Society 41, 70±77

206. Gould D. (1996) Using vignettes to collect data for nursing research studies: how valid are the fndings? Journal of Clinical Nursing 5,207±212.

207. Meurier CE, Vincent CA, Parmar DG. Nurses' responses to severity dependent errors: a study of the causal attributions made by nurses following an error. J Adv Nurs 1998;27:349–54

208. McKinstry B. (2000) Do patients wish to be involved in decisionmaking in the consultation? A cross sectional survey with videovignettes. British Medical Journal 321, 867±871.

209. Atzmüller, C. , & Steiner, P. M. (2010). Experimental vignette studies in survey research. Methodology, 6(3), 128–138.

210. Hughes R, Huby M. The application of vignettes in social and nursing research. J Adv Nurs. 2002 Feb;37(4):382-6.

211. Leistikow I, Mulder S, Vesseur J, et al. Learning from incidents in healthcare: the journey, not the arrival, matters. BMJ Quality & Safety 2017;26:252-256.

212. Waterson P, Catchpole Khuman factors in healthcare: welcome progress, but still scratching the surface. BMJ Quality & Safety 2016;25:480-484.

213. Waterson P. Promoting systemic incident analysis in healthcare—key challenges and ways forwards, *International Journal for Quality in Health Care*, Volume 33, Issue 4, 2021

214. Rowe G, Wright G. The Delphi technique as a forecasting tool: Issues and analysis. Int. J. Forecasting, 15 (4) (1999), pp. 353-355

215. Stewart, J. (2001). Is the Delphi technique a qualitative method? Medical Education, 35, 922-923.

216. Stitt-Gohdes W, Crews TB. Journal of Career and Technical Education, Vol. 20, No. 2, Spring, 2004

217. Armitage G , Moore S , Reynolds C , et al . Patient-reported safety incidents as a new source of patient safety data: an exploratory comparative study in an acute hospital in England. J Health Serv Res Policy 2017; 20:135581961772756

218. O'Hara JK, Reynolds C, Moore S, et al. What can patients tell us about the quality and safety of hospital care? Findings from a UK multicentre survey study. BMJ Quality & Safety 2018;27:673-682.

219. Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications*. London: Addison-Wesley.

220. Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R; Global Consensus Group. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am J Gastroenterol. 2006 Aug;101(8):1900-20; quiz 1943.

221. Bennett C et al. Consensus statements for management of Barrett's dysplasia and early-stage esophageal adenocarcinoma, based on a Delphi process. Gastroenterology. 2012 Aug;143(2):336-46.

222. Eubank, B.H., Mohtadi, N.G., Lafave, M.R. et al. Using the modified Delphi method to establish clinical consensus for the diagnosis and treatment of patients with rotator cuff pathology. BMC Med Res Methodol 16, 56 (2016).

223. Vanhaecht K, et al; European Researchers' Network Working on Second Victims (ERNST). An Evidence and Consensus-Based Definition of Second Victim: A Strategic Topic in Healthcare Quality, Patient Safety, Person-Centeredness and Human Resource Management. Int J Environ Res Public Health. 2022 Dec 15;19(24):16869.

224. Ockenden D. Ockenden report: findings, conclusions and essential actions from the independent review of the maternity services at the Shrewsbury and Telford Hospital NHS Trust. 2022. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1064302/Final-Ockenden-Report-web-accessible.pdf. Accessed November 1, 2022.

225. Infected Blood Inquiry. 2022. Available at: https://www.infectedbloodinquiry.org.uk/. Accessed November 1, 2022.

226. Carson-Stevens A, Hibbert P, Williams H, et al. Characterising the nature of primary care patient safety incident reports in the England and Wales National Reporting and Learning System: a mixed-methods agenda-setting study for general practice. *Health Serv Deliv Res*. 2016;4.

227. Gens-Barbera M, Hernández-Vidal N, Vidal-Esteve E, et al. Analysis of patient safety incidents in primary care reported in an

electronic registry application. *Int J Environ Res Public Health*. 2021;18:8941.

228. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169:467–473.

229. Harrison R, Jones B, Gardner P, et al. Quality Assessment with Diverse Studies (QuADS): an appraisal tool for methodological and reporting quality in systematic reviews of mixed- or multi-method studies. *BMC Health Serv Res*. 2021;21:144.

230. Sirriyeh R, Lawton R, Gardner P, et al. Reviewing studies with diverse designs: the development and evaluation of a new tool. J Eval Clin Pract. 2012;18:746–752.

231. Bengtsson M. How to plan and perform a qualitative study using content analysis. *NursingPlus Open*. 2016;2:8–14.

232. Robbins T, Tipper S, King J, et al. Evaluation of learning teams versus root cause analysis for incident investigation in a large United Kingdom National Health Service Hospital. *J Patient Saf*. 2021;17:e1800–e1805.

233. van der Starre C, van Dijk M, van den Bos A, et al. Paediatric critical incident analysis: lessons learnt on analysis, recommendations and implementation. *Eur J Pediatr*. 2014;173:1449–1457.

234. Hettinger AZ, Fairbanks RJ, Hegde S, et al. An evidence-based toolkit for the development of effective and sustainable root cause analysis system safety solutions. *J Healthc Risk Manag*. 2013;33:11–20.

235. Cafazzo JA, St-Cyr O. From discovery to design: the evolution of human factors in healthcare. *Healthc Q*. 2012;15 Spec No:24–29.

236. Wood LJ, Wiegmann DA. Beyond the corrective action hierarchy: a systems approach to organizational change. Int J Qual Health Care. 2020;32:438–444.

237. Liberati EG, Peerally MF, Dixon-Woods M. Learning from high risk industries may not be straightforward: a qualitative study of the

hierarchy of risk controls approach in healthcare. *Int J Qual Health Care*. 2018;30:39–43.

238. Mills PD, Neily J, Luan D, et al. Using aggregate root cause analysis to reduce falls. Jt Comm J Qual Patient Saf. 2005;31:21–31.

239. Mills PD, Neily J, Luan D, et al. Actions and implementation strategies to reduce suicidal events in the Veterans Health Administration. Jt Comm J Qual Patient Saf. 2006;32:130–141.

240. Mills PD, Neily J, Kinney LM, et al. Effective interventions and implementation strategies to reduce adverse drug events in the Veterans Affairs (VA) system. Qual Saf Health Care. 2008;17:37–46.

241. Mills PD, Huber SJ, Vince Watts B, et al. Systemic vulnerabilities to suicide among veterans from the Iraq and Afghanistan conflicts: review of case reports from a national Veterans Affairs database. Suicide Life Threat Behav. 2011;41:21–32.

242. Mills PD, Gallimore BI, Watts BV, et al. Suicide attempts and completions in Veterans Affairs nursing home care units and long-term care facilities: a review of root-cause analysis reports. Int J Geriatr Psychiatry. 2016;31:518–525.

243. Charuluxananan S, Suraseranivongse S, Jantorn P, et al. Multicentered study of model of anesthesia related adverse events in Thailand by incident report (The Thai Anesthesia Incidents Monitoring Study): results. J Med Assoc Thai. 2008;91:1011–1019.

244. Hooker AB, Etman A, Westra M, et al. Aggregate analysis of sentinel events as a strategic tool in safety management can contribute to the improvement of healthcare safety. Int J Qual Health Care. 2019;31:110–116.

245. Neily J, Silla ES, Sum-Ping SJT, et al. Anesthesia adverse events voluntarily reported in the Veterans Health Administration and lessons learned. Anesth Analg. 2018;126:471–477.

246. Igene OO, Johnson C. Analysis of medication dosing error related to Computerised Provider Order Entry system: a comparison of ECF,

HFACS, STAMP and AcciMap approaches. *Health Informatics J*. 2019;1–26.

247. Igene OO, Johnson CW, Long J. An evaluation of the formalised AcciMap approach for accident analysis in healthcare. *Cogn Tech Work*. 2022;24:161–181.

248. Patriarca R, Gravio GD, Woltjer R, et al. Framing the FRAM: a literature review on the functional resonance analysis method. *Saf Sci*. 2020;129.

249. Anjalee JAL, Rutter V, Samaranayake NR. Application of failure mode and effects analysis (FMEA) to improve medication safety in the dispensing process—a study at a teaching hospital, Sri Lanka. *BMC Public Health*. 2021;21:1430.

250. Slade JJ, Wrzesniewski CE, Hunter OO, et al. Complementing root cause analysis with improvement strategies to optimize venous thromboembolism prophylaxis in patients with epidural catheters. *Qual Manag Health Care*. 2020;29:253–259.

251. VHA National Center for Patient Safety. Guide to Performing a Root Cause Analysis. 2021. Available at: https://www.patientsafety.va.gov/docs/RCA-Guidebook_02052021.pdf. Accessed March 28, 2022.

252. Manuele FA. Achieving risk reduction, effectively. *Proc Saf Environ Protect*. 2006;84(B3):184–190.

253. French SD, Green SE, O'Connor DA, et al. Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. *Implement Sci*. 2012;7:38.

254. Lea W, Lawton R, Vincent C, O'Hara J. Exploring the 'Black Box' of Recommendation Generation in Local Health Care Incident Investigations: A Scoping Review. J Patient Saf 2023;19:553–63.

255. Peerally, MF, Carr S, Waring J, Martin G, Dixon-Woods Mary. Risk Controls Identified in Action Plans Following Serious Incident

Investigations in Secondary Care: A Qualitative Study. Journal of Patient Safety 20(6):p 440-447, September 2024.

256. Wailling J, Kooijman A, Hughes J, O'Hara JK. Humanizing harm: using a restorative approach to heal and learn from adverse events. *Health Expect*. 2022; 25: 1192-1199.

257. Suchman H, Presser S. Questions and answers in attitude surveys: Experiments on question form, wording, and contexts. New York: Academic Press, 1981.

258. Boxebeld S. Ordering effects in discrete choice experiments: A systematic literature review across domains. Journal of Choice Modelling 2024;51:100489.

259. R Core Team. R foundation for statistical computing. R: A language and environment for statistical computing; 2021. Available: https://www.R-project.org

260. Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67.

261. Hess TM, McGee KA, Woodburn SM, et al. Age-related priming effects in social judgments. Psychol Aging 1998;13:127–37.

262. Chen Y, Blanchard-Fields F. Unwanted thought: age differences in the correction of social judgements. Psychol Aging 2000;15:475–82.

263. Strömwall LA, Landström S, Alfredsson H. Perpetrator characteristics and blame attributions in a stranger rape situation. EJPALC 2014;6:63–7.

264. Morgenroth T, Ryan M. Gender in a social psychology context. Oxford Research Encyclopedia of Psychology; 2018. Available: https://oxfordre.com/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-309

265. Atari M, Lai MHC, Dehghani M. Sex differences in moral judgements across 67 countries. Proc Biol Sci 2020;287:20201201.

266. Dekker SWA. Reconstructing human contributions to accidents: the new view on error and performance. J Safety Res 2002;33:371–85.

267. Christensen-Szalanski JJJ, Willham CF. The hindsight bias: A meta-analysis. Organ Behav Hum Decis Process 1991;48:147–68.

268. Fischhoff B, Beyth R. 'I knew it would happen'— Remembered probabilities of once-future things. Organizational Behavior & Human Performance 1975;13:1–16.

269. Hawkins SA, Hastie R. Hindsight: Biased judgments of past events after the outcomes are known. Psychol Bull 1990;107:311–27.

270. Gerken M. Assessing the Evidence for Outcome Bias and Hindsight Bias. RevPhilPsych 2024;15:237–52.

271. Henriksen K, Kaplan H. Hindsight bias, outcome knowledge and adaptive learning. Qual Saf Health Care 2003;12 Suppl 2:ii46–50.

272. MacLean CL. Cognitive bias in workplace investigation: Problems, perspectives and proposed solutions. Appl Ergon 2022;105:103860.

273. Meyer CO. Can one 'prove' that a harmful event was preventable? Conceptualizing and addressing epistemological puzzles in post incident reviews and investigations. Risk Hazard & Crisis Pub Pol 2024;15:374–92.

274. Patient safety incident response framework. NHS England; 2024. Available: https://www.england.nhs.uk/long-read/patient-safety-incident-response-framework/ [Accessed 16 Aug 2024].

275. Milligan C, Allin S, Farr M, et al. Mandatory reporting legislation in Canada: improving systems for patient safety? Health Econ Policy Law 2021;16:355–70.

276. Adams GS, Converse BA, Hales AH, et al. People systematically overlook subtractive changes. Nature New Biol 2021;592:258–61.

277. Rae AJ, Provan DJ, Weber DE, et al. Safety clutter: the accumulation and persistence of ‘safety’ work that does not contribute to operational safety. PPHS 2018;16:194–211.

278. Halligan D, Janes G, Conner M, et al. Identifying Safety Practices Perceived as Low Value: An Exploratory Survey of Healthcare Staff in the United Kingdom and Australia. J Patient Saf 2023;19:143–50.

279. Catino M. Blame culture and defensive medicine. Cogn Tech Work 2009;11:245–53.

280. Parker J, Davies B. No Blame No Gain? From a No Blame Culture to a Responsibility Culture in Medicine. J Appl Philos 2020;37:646–60.

281. Stanovich KE, West RF. On the relative independence of thinking biases and cognitive ability. J Pers Soc Psychol 2008;94:672–95.

282. West RF, Meserve RJ, Stanovich KE. Cognitive sophistication does not attenuate the bias blind spot. J Pers Soc Psychol 2012;103:506–19.

283. Hutchinson A, Coster JE, Cooper KL, et al. A structured judgement method to enhance mortality case note review: development and evaluation. BMJ Qual Saf 2013;22:1032–40.

284. Janis IL. Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes. Houghton Mifflin, 1972.

285. Stasser G, Titus W. Pooling of unshared information in group decision making: Biased information sampling during discussion. J Pers Soc Psychol 1985;48:1467–78.

286. Greitemeyer T, Schulz-Hardt S. Preference-consistent evaluation of information in the hidden profile paradigm: beyond group-level explanations for the dominance of shared information in group decisions. J Pers Soc Psychol 2003;84:322–39.

287. Chiu C, Sharon S-L, Evelyn WMAu. Culture and social cognition. In: Carlston D, ed. The oxford handbook of social cognition. Oxford Academic, 2013.

288. Benner L. Rating Accident Models and Investigation Methodologies. Journal of Safety Research. 1985, Vol. 16 pp. 105-126

289. Kirwan B. Human error identification techniques for risk assessment of high risk systems- part 1: review and evaluation of techniques. Applied ergonomics 1998 29 (3), 157-177

290. Dalkey N, Helmer O. An experimental application of the Delphi method to the use of experts. Manag Sci. 1963;9(3):458–67.

291. Dalkey N. An experimental study of group opinion: The Delphi method. Futures. Volume 1, Issue 5, September 1969, Pages 408-426

292. Boulkedid et al Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. PLoS ONE 2011 6(6): e20476

293. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. Journal of Advanced Nursing, 2000, 32(4), 1008-1015

294. Woodcock T, Adeleke Y, Goeschel C. et al. A modified Delphi study to identify the features of high quality measurement plans for healthcare improvement projects. BMC Med Res Methodol 20, 8 (2020).

295. Auger, C.P., ed. (1975). Use of Reports Literature. London: Butterworth. ISBN 978-0-408-70666-7.

296. Bos K, Dongelmans DA, Groeneweg J, Legemate DA, Leistikow IP, van der Laan MJ. Criteria for recommendations after perioperative sentinel events. BMJ Open Qual. 2021 Sep;10(3):e001493.

297. NHS Education for Scotland. Learning Response Review and Improvement Tool [internet]. 2025. Available online from https://hssib-ovd42x6f-media.s3.amazonaws.com/production-assets/documents/Learning_Response_Review_and_Improvement_Tool_accessbile_PDF_v2.0_July_2025.pdf

298. de Kam D, Kok J, Grit K, Leistikow I, Vlemminx M, Bal R. How incident reporting systems can stimulate social and participative learning: A mixed-methods study. Health Policy. 2020 Aug;124(8):834-841.

299. Anderson JE, Watt AJ. Using Safety-II and resilient healthcare principles to learn from Never Events. Int J Qual Health Care. 2020 May 20;32(3):196-203.

300. NHS England. Patient safety incident response standards [internet]. 2024. Last accessed 29th July 2025. Available from https://www.england.nhs.uk/long-read/patient-safety-incident-response-standards/

301. Pham JC, Hoffman C, Popescu IC, Ijagbemi OM, Canadian Patient Safety Institute. Concise Incident Analysis Method Pilot Study. 2014.

Available online from https://era.library.ualberta.ca/items/69c7ae45-1a1c-441b-918b-65538254dbe3

302. Fischhoff B. Hindsight: thinking backward? Oregon Res Inst Res Monogr 1974;14: No. 1.

303. Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). Behind human error: cognitive systems, computers and hindsight. Dayton, OH: CSERIAC.

304. Morita, P. P., Burns, C. M., & Calil, S. J. (2009). The Influence of Strong Recommendations, Good Incident Reports and a Monitoring System over an Incident Investigation System for Healthcare Facilities. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 53(22), 1679-1683.

305. Zelik, D.J., Patterson, E.S., Woods, D.D., 2009. Measuring attributes of rigor in information analysis. In: Patterson, E.S., Miller, J. (Eds.), Macrocognition Metrics and Scenarios: Design and Evaluation for Real-world Teams. Ashgate, Aldershot, UK

306. Tomić B, Spacojević Brkić V. EFFECTIVE ROOT CAUSE ANALYSIS AND CORRECTIVE ACTION PROCESS. JOURNAL OF ENGINEERING MANAGEMENT AND COMPETITIVENESS Vol. 1, No. 1/2, 2011, 16-20

307. Hollnagel, E., & Leonhardt, J. (2014). From Safety-I to Safety-II: A White Paper. Eurocontrol. https://www.eurocontrol.int/sites/default/files/content/documents/nm/safety/safety_whitepaper_sept_2013-web.pdf

308. Underwood P, Waterson P. Systemic accident analysis: examining the gap between research and practice. Accid Anal Prev. 2013 Jun;55:154-64.

309. Drupsteen, L., & Guldenmund, FW. (2014). What is learning? A review of the safety literature to define learning from incidents, accidents and disasters. Journal of Contingencies and Crisis Management, 22(2), 81-96. https://doi.org/10.1111/1468-5973.12039

310. Hollnagel E. Safety-I and Safety-II: the past and future of safety management Aldershot, UK: Ashgate; 2014.

311. Hopkins A. Issues in safety science. Safety Science Volume 67, August 2014, Pages 6-14

312. Anderson JE, Kodate N. Learning from patient safety incidents in incident review meetings: Organisational factors and indicators of analytic process effectiveness. Safety Science Volume 80, December 2015, Pages 105-114

313. Stavropoulou C, Doherty C, Tosey P. How Effective Are Incident-Reporting Systems for Improving Patient Safety? A Systematic Literature Review. The Milbank Quarterly 2015 Volume 93, Issue 4 (pages 826–866)

314. Vincent C, Carthey J, Macrae C, Amalberti R. Safety analysis over time: seven major changes to adverse event investigation. Implement Sci. 2017 Dec 28;12(1):151.

315. Smith AF, Plunkett E. People, systems and safety: resilience and excellence in healthcare practice. Anaesthesia. 2019 Apr;74(4):508-517.

316. Stemn, E., Hassall, M.E., Bofinger, C. et al. Assessing the Quality of Incident Investigations and its Effect on Safety Performance: a Study of the Ghanaian Mining Industry. Mining, Metallurgy & Exploration 36, 1145–1154 (2019).

317. Bos K, Dongelmans DA, Greuters S, Kamps GJ, van der Laan MJ. The next step in learning from sentinel events in healthcare. BMJ Open Qual. 2020 Feb;9(1):e000739.

318. Wiegmann DA, Wood LJ, Solomon DB, Shappell SA. Implementing a human factors approach to RCA2 : Tools, processes and strategies. J Healthc Risk Manag. 2021 Jul;41(1):31-46.

319. Solomon, D., Wood, L., & Wiegmann, D. (2021). Implementing A Human Factors Approach To Rca2: Analysis Of Resultant Safety Recommendations. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 65(1), 628-632.

320. Thallapureddy S, Sherratt F, Bhandari S, Hallowell M, Hansen H. Exploring bias in incident investigations: An empirical examination using construction case studies. J Safety Res. 2023 Sep;86:336-345.

321. Hutchinson B, Dekker S, Rae A. How audits fail according to accident investigations: A counterfactual logic analysis. Process Saf Prog. 2024; 43(3): 441-454.

322. Lamé G, Liberati EG, Canham A, Burt J, Hinton L, Draycott T, Winter C, Dakin FH, Richards N, Miller L, Willars J, Dixon-Woods M. Why is safety in intrapartum electronic fetal monitoring so hard? A qualitative study combining human factors/ergonomics and social science analysis. BMJ Qual Saf. 2024 Mar 25;33(4):246-256.

323. Sujan M, Lounsbury O, Pickup L, et al., (2024) What kinds of insights do Safety-I and Safety-II approaches provide? a critical reflection on the use of SHERPA and FRAM in healthcare. Safety Science, Volume 173, May 2024, Article Number 106450

324. Carthey J, Clinical Human Factors Group. Implementing Human Factors in healthcare 'How to' guide - volume 2 'Taking further steps'. 2013 Clinical Human Factors Group. Available online from https://chfg.org/how-to-guide-to-human-factors-volume-2/

325. Lee 2004. Topics in Patient Safety: How to Make the Most of Actions and Outcome Measures. VA National Center for Patient Safety

326. Expert Advisory Group. Report of the Expert Advisory Group: Healthcare Safety Investigation Branch. 2016. Available from https://assets.publishing.service.gov.uk/media/5a7f6157e5274a2e87db580f/hsibreport.pdf

327. NHS Improvement. The future of NHS patient safety investigation: engagement feedback. 2018. No longer available online.

328. O'Hara J, Ramsey L, McHugh S, Langley J, Waring J, Simms-Ellis R, et al. Co-designing and testing the learn together guidance to support patient and family involvement in patient safety investigations: a mixed-methods study. Health Soc Care Deliv Res 2025;13(18).

329. Peerally, Farhad M. (2021). Improving risk controls following root cause analysis of serious incidents in healthcare. University of Leicester. Thesis. https://doi.org/10.25392/leicester.data.15112434.v1

330. Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of boot strap data expansion. BMC Medical Research Methodology 2005, 5:37

331. Hsu C, Sandford BA. Minimizing non-response in the Delphi process: how to respond to non-response. Pract Assess Res Eval 2007;12

332. Taylor RM, Feltbower RG, Aslam N, Raine R, Whelan JS, Gibson F. Modified international e-Delphi survey to define healthcare professional competencies for working with teenagers and young adults with cancer. BMJ Open. 2016 May 3;6(5):e011361.

333. McHugh, S., Louch, G., Ludwin, K. et al. Involvement in serious incident investigations: a qualitative documentary analysis of NHS trust policies in England. BMC Health Serv Res 24, 1207 (2024).

334. Kaboli PJ , Glasgow JM , Jaipaul CK , et al . Identifying medication misadventures: poor agreement among medical record, physician, nurse, and patient reports. Pharmacotherapy 2010;30:529–38.

335. Armitage G , Moore S , Reynolds C , et al . Patient-reported safety incidents as a new source of patient safety data: an exploratory comparative study in an acute hospital in England. J Health Serv Res Policy 2017;20:135581961772756.

336. O'Hara JK, Reynolds C, Moore S, et al. What can patients tell us about the quality and safety of hospital care? Findings from a UK multicentre survey study. BMJ Quality & Safety 2018;27:673-682.

337. Brennan PF, Safran C. "Patient safety: Remember who it's really for." *International journal of medical informatics* 73 7-8 (2004): 547-50 .

338. Wu AW, Shapiro J, Harrison R, et al. The Impact of Adverse Events on Clinicians: What's in a Name? J Patient Saf 2020;16:65–72

339. Quillivan RR, Burlison JD, Browne EK, et al. Patient Safety Culture and the Second Victim Phenomenon: Connecting Culture to Staff Distress in Nurses. Jt Comm J Qual Patient Saf 2016;42:377–86.

340. van Marum S, Verhoeven D, de Rooy D. The Barriers and Enhancers to Trust in a Just Culture in Hospital Settings: A Systematic Review. J Patient Saf 2022;18:e1067–75.

341. Turner K, Sveticic J, Grice D, et al. Restorative just culture significantly improves stakeholder inclusion, second victim experiences and quality of recommendations in incident responses. JHA 2022;11:8.

342. Kaur M, De Boer RJ, Oates A, et al. Restorative Just Culture: a Study of the Practical and Economic Effects of Implementing Restorative Justice in an NHS Trust. MATEC Web Conf 2019;273:01007.

343. Kim S-A, Kim E-M, Lee J-R. Causes of Nurses' Second Victim Distress: An Objective Analysis. Qual Manag Health Care 2022;31:122–9.

344. Murray JS, Clifford J, Larson S, et al. Implementing Just Culture to Improve Patient Safety. Mil Med 2023;188:1596–9.

345. Dixon-Woods M, McNicol S, Martin G. Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature. BMJ Quality & Safety 2012;21:876-884.

346. Page B, Irving D, Amalberti R, et al. Health services under pressure: a scoping review and development of a taxonomy of adaptive strategies. BMJ Quality & Safety 2024;33:738-747.

347. Leviton LC, Melichar L. Balancing stakeholder needs in the evaluation of healthcare quality improvement. BMJ Quality & Safety 2016;25:803-807.

348. Williams VW, Finkelstein JB. Speaking and listening: The importance of stakeholder engagement in quality improvement in pediatric urology. J Pediatr Urol. 2023 Dec;19(6):792-799.

349. Lea W, Budworth L, O'Hara J, Vincent C, Lawton R. Investigators are human too: outcome bias and perceptions of individual culpability in patient safety incident investigations. BMJ Quality & Safety Published Online First: 10 February 2025.

350. Mahase E. Invest in health workforce or risk collapse, WHO warns governments. BMJ 2023;380:713

351. Joseph O, Mir G, Fylan B, Essler P, Lawton R. Incivility experiences of racially minoritised hospital staff, consequences for them and implications for patient care: An international scoping review. Sociol Health Illn. 2025 Jan;47(1):e13760.

352. Aunger J, Maben J, Westbrook JI. How unprofessional behaviours between healthcare staff threaten patient care and safety. Expert Rev Pharmacoecon Outcomes Res. 2025 Jan 30:1-4.

353. Percarpio KB, Watts BV, Weeks WB. The effectiveness of root cause analysis: what does the literature tell us? Jt Comm J Qual Patient Saf 2008;34:391–8.

354. Martens J, Van Gerven E, Lannoy K, et al. Serious reportable events within the inpatient mental health care: Impact on physicians and nurses. Rev Calid Asist 2016;31 Suppl 2:26–33.

355. Castro Lopes S, Guerra-Arias M, Buchan J, et al. A rapid review of the rate of attrition from the health workforce. Hum Resour Health 2017;15:21.

356. Ratwatte M. The NHS workforce crisis is a retention crisis. BMJ 2023;380:602.

357. Clarkson J, Dean J, Ward J, Komashie A, Bashford T. A systems approach to healthcare: from thinking to -practice. Future Healthc J. 2018 Oct;5(3):151-155.

358. Johnston, N., 1995. Spring. Do Blame and Punishment have A Role in Organizational Risk Management? Flight Deck, pp. 33–36.

359. Shojania KG, Dixon-Woods M. 'Bad apples': time to redefine as a type of systems problem? BMJ Qual Saf. 2013 Jul;22(7):528-531.

360. Health Services Safety Investigations Body (HSSIB). 2025 Who we are. Available at: https://www.hssib.org.uk/about-us/who-we-are/ (Accessed 7 May 2025)

361. Dror, I. E. (2016). A hierarchy of expert performance. *Journal of Applied Research in Memory and Cognition, 5*(2), 121–127.

362. Woodcock K, Smiley A. Factors affecting the classes of causal explanations by industrial safety specialists. ADVANCES IN OCCUPATIONAL ERGONOMICS AND SAFETY 1997 Jun 1:69-72.

363. Morris, M.W., Moore, P.C., Sim, D.H., 1999. Choosing remedies after accidents: counterfactual thoughts and the focus on fixing 'human error. Psychonomic Bull. Rev. 6, 579–585.

364. Woodcock K, Drury CG, Smiley A, Ma J. 2005. Using simulated investigations for accident investigation studies. Appl. Ergon. 36 (1), 1–12.

365. Rosset, E., 2008. It's no accident: our bias for intentional explanations. Cognition: Int. J. Cognit. Sci. 108 (3), 771.

366. Roese, N.J., Vohs, K.D., 2012. Hindsight bias. Perspect. Psychol. Sci. 7 (5), 411–426.

367. Dekker, S.W.A., 2004. The hindsight bias is not a bias and not about history. Hum. Factors Aero Saf. 4 (2), 87–99.

368. Machen S. Thematic reviews of patient safety incidents as a tool for systems thinking: a quality improvement report: BMJ Open Quality 2023;12:e002020.

369. Macrae C. Investigating for improvement? Five strategies to ensure national patient safety investigations improve patient safety. *Journal of the Royal Society of Medicine*. 2019;112(9):365-369.

370. Health Equity Assessment Tools (2025) MNSI. Available at: https://www.mnsi.org.uk/for-nhs/health-equity-assessment-and-health-equity-warning-score/ (Accessed: 19 July 2025).

371. Greenhalgh T, Russell J. Reframing evidence synthesis as rhetorical action in the policy making drama. Healthc Policy. 2006;1(2):34.

372. Jones L, Exworthy M. Framing in policy processes: a case study from hospital planning in the National Health Service in England. Soc Sci Med. 2015;124:196–204.

373. Rollenhagen C. The context and habits of accident investigation practices: A study of 108 Swedish investigators. Safety Science. 2010

374. Beijaard, D., Verloop, N. and Vermunt, J. (2004). Reconsidering research on teachers' professional identity. Teaching and Teacher Education. Vol. 20 (2): 107-128.

375. Neary, S. (2014). Professional identity: What I call myself defines who I am. Career Matters, 2(3), 14–15. https://core.ac.uk/download/pdf/46170813.pdf

376. Fitzgerald, A. (2020). Professional identity: A concept analysis. Nursing Forum, 55(3), 447–472.

377. Savin-Baden, M., & Major, C. (2013). Qualitative research: The essential guide to theory and practice. Routledge.

378. Wilson C, Janes G, Williams J. Identity, positionality and reflexivity: relevance and application to research paramedics. Br Paramed J. 2022 Sep 1;7(2):43-49.

379. Gurr H, Oliver L, Harvey O, Subedi M, Teijlingen E. The Importance of Positionality for Qualitative Researchers. Dhaulagiri Journal of Sociology and Anthropology Vol. 18(1), 2024, pp. 48-54.

380. Mclaughlin S. When I say … positionality. Med Educ. 2024 Oct;58(10):1149-1150.

381. Sikes P. "Methodology, Procedures, and Ethical Concerns." In Doing Educational Research: A Guide for First-Time Researchers, ed. Clive Opie, 15–33. 2004. London: SAGE Publications.

382. Jacobson, D., & Mustafa, N. Social Identity Map: A Reflexivity Tool for Practicing Explicit Positionality in Critical Qualitative Research. International Journal of Qualitative Methods, 18, 2019.

383. Holmes A. Researcher positionality–A consideration of its influence and place in qualitative research–A new researcher guide. Shanlax International Journal of Education, 2020, 8(4), 1–10.

Appendix 1: Fictional Incident to demonstrate application of the OAM

Incident Summary: The incident takes place on a hospital ward, a nurse (Alex) on their drug round notices that the next patient (Sam) has been prescribed new antibiotics to treat a hospital acquired pneumonia. The antibiotics are to be given intravenously. The nurse prepares and administers the antibiotics. A short time later the patient becomes unwell. Despite the efforts of the resuscitation team (made up of doctors and nurses) the patient dies. It is likely they died because of anaphylactic shock triggered by the intravenous antibiotics. The antibiotics given were from the penicillin class.

One of the resuscitation team notices, on the electronic prescribing system, that a penicillin allergy is recorded for Sam. An incident report is submitted via the online system.

Initial fact finding: The ward manager speaks to Alex about the incident. Alex says they were not aware of the allergy, and that the patient did not have a red allergy wrist band on. They said that the antibiotic had been prescribed by the doctor and so they assumed that it was safe to give and that the doctor would have checked the allergies. There is a yellow box indicating allergies on the electronic prescribing system display, but you have to click on it to view what the allergies the patient has. The ward manager also spoke to the doctor who

had prescribed the antibiotic (Sidney), who said they had not been aware that the patient had any allergies.

Initial analysis: Active failures are unsafe acts committed by people;(87Reason 2000) in this case Sidney prescribed and Alex administered a penicillin antibiotic to a patient with a known penicillin allergy. It is not uncommon for analysis and investigation to look little further than the often obvious active failures of staff in relation to an incident.(87Reason 2000; 144Lawton 2012; 145Peerally 2022) It is the purpose of the systems approach and OAM to encourage exploration beyond the point of an active failures.(87Reason 2000) The following paragraphs will explore selected defences/barriers and contributory factors in this incident in order illustrate the use of AOM rather than represent a complete investigation.

Analysis of defences/barriers: The hospital in which this incident occurred has a policy of placing a red plastic bracelet on patients' wrists to highlight known allergies. This is an example of a 'defence' and acts as a safeguard to mitigate the consequence of human failure.(87Reason 2000) The patient in this incident had been given a plastic bracelet but were suffering from acute confusion; they had been agitated during the night and pulled off this bracelet. The bracelet had not been replaced before Alex had come to administer the antibiotics. The defences illustrated in Figure 4 have holes representing the fact that the defences may or may not be intact; ideally defences are totally effective and never fail, but in reality, this is unlikely. As in this incident, the red allergy bracelet defence, is only effective if it is placed on the patient's wrist, and even then needs to be seen and understood by staff to have a chance of preventing harm. There were a number of other potential defences or barriers in this incident such as allergy alerts on the electronic prescribing system and practice of asking a patient if they have any allergies before administering medication.

Analysis of contributory Factors: Selected examples of contributory factors are discussed here in order to illustrate the OAM (Figure 3) within this fictional incident; a full investigation would likely uncover much more detail.

Patient factors: The patient in this incident was acutely confused and agitated. They had a mixed delirium, meaning that at times they were agitated and other times drowsy. They had removed their own allergy bracelet during the night, were not able to report their allergy to the nurse, and there was a delay in the detection of their clinical deterioration due to cognitive impairment.

Work Environmental Factors: The ward manager was aware that on the day this incident occurred there were some issues with staffing. There were less nurses than there should be, and one of the nurses was a locum nurse who was not as familiar with this ward or hospital. As a result Alex was caring for more patients than usual and having to provide support to the other staff. Alex was having to frequently interrupt their drug round to provide support to other staff.

Electronic Information Systems and Technology: The electronic prescribing system is designed in such a way that 1) the doctor was able to prescribe a penicillin antibiotic to a patient with known and recorded penicillin allergy, 2) there was no allergy warning when the nurse administered the antibiotic to the patient.

Organisational, Management and Cultural Factors: Due to financial pressures the hospital made changes to how they paid nurses to work extra shifts; resulting in less pay than previously. This meant that nurses who usually worked in the hospital, and who were familiar with the ward, were less likely to take on extra work and gaps in staffing were filled by locum nurses.

The purpose of presenting this fictional incident is to illustrate how the OAM can be used to explore the multiple contributory factors behind an incident. It is important to highlight that there are two immediate active failures that were identified (prescription and then administration of the antibiotic), each of which

may have their own set of contributory factors, and that it may be that beyond the harm caused by administration of the antibiotic there may have been a delay in recognition of the patients deterioration; investigation of which may reveal further contributory factors.

Appendix 2: Scoping review search terms

Embase:	Medline:	PyschInfo	CINAHL
1. incident*.tw.	1. incident*.tw.	1. incident*.tw.	1. TI incident* OR AB incident*
2. patient safety.tw.	2. patient safety.tw.	2. patient safety.tw.	2. TI patient safety OR AB patient safety
3. exp patient safety/	3. exp patient safety/	3. exp patient safety/	3. (MH "Patient Safety+")
4. adverse event*.tw.	4. adverse event*.tw.	4. adverse event*.tw.	4. TI adverse event* OR AB adverse event*
5. near miss*.tw.	5. near miss*.tw.	5. near miss*.tw.	5. TI near miss* OR AB near miss*
6. exp medical error/	6. exp medical errors/	6. medical error*.tw.	6. (MH "Health Care Errors+")
7. sentinel event*.tw.	7. sentinel event*.tw.	7. sentinel event*.tw.	7. TI sentinel event* OR AB sentinel event*
8. medical error*.tw.	8. medical error*.tw.	8. exp Errors/	8. TI medical error* OR AB medical error*
9. exp medication error/	9. exp medication errors/		9. (MH "Medication Errors+")
10. investigation*.tw.	10. investigation*.tw.	9. investigation*.tw.	10. TI investigation* OR AB investigation*
11. exp "root cause analysis"/	11. exp "root cause analysis"/	10. exp Causal Analysis/	11. (MM "Root Cause Analysis")
12. (analyz* adj4 incident*).tw.	12. (analyz* adj4 incident*).tw.	11. (analyz* adj4 incident*).tw.	12. TI analyz* N4 incident* OR AB analyz* N4 incident*
13. (analys* adj4 event*).tw.	13. (analys* adj4 event*).tw.	12. (analys* adj4 event*).tw.	13. TI analys* N4 event* OR AB analys* N4 event*
14. (analyz* adj4 event*).tw.	14. (analyz* adj4 event*).tw.	13. (analyz* adj4 event*).tw.	14. TI analyz* N4 event* OR AB analyz* N4 event*
15. (analys* adj4 incident*).tw.	15. (analys* adj4 incident*).tw.	14. (analys* adj4 incident*).tw.	15. TI analys* N4 incident* OR AB analys* N4 incident*
16. "root Cause analys*".tw.	16. "root Cause analys*".tw.	15. "root Cause analys*".tw.	16. TI "root Cause analys*" OR AB "root Cause analys*"
17. Risk control*.tw.	17. Risk control*.tw.	16. Risk control*.tw.	17. TI Risk control* OR AB Risk control*
18. Recommendation*.tw.	18. Recommendation*.tw.	17. Recommendation*.tw.	18. TI Recommendation* OR AB Recommendation*
19. Learn*.tw.	19. Learn*.tw.	18. Learn*.tw.	19. TI Learn* OR AB Learn*
20. exp risk management/	20. exp risk management/	19. exp Risk Management/	20. (MH "Risk Management+")
21. lesson*.tw.	21. lesson*.tw.	20. lesson*.tw.	21. TI lesson* OR AB lesson*
22. action*.tw.	22. action*.tw.	21. action*.tw.	22. TI action* OR AB action*
23. prevent*.tw.	23. prevent*.tw.	22. prevent*.tw.	23. TI prevent* OR AB prevent*
24. response*.tw.	24. response*.tw.	23. response*.tw.	24. TI response* OR AB response*
25. improvement*.tw.	25. improvement*.tw.	24. improvement*.tw.	25. TI improvement* OR AB improvement*
26. exp hospital/	26. exp hospitals/	25. exp HOSPITALS/	26. (MH "Hospitals+")
27. hospital*.tw.	27. hospital*.tw.	26. hospital*.tw.	27. TI hospital* OR AB hospital*
28. secondary care.tw.	28. secondary care.tw.	27. secondary care.tw.	28. TI secondary care OR AB secondary care
29. exp health care/	29. exp "Delivery of Health Care/	28. exp Health Care Services/	29. (MM "Tertiary Health Care") OR (MM "Secondary Health Care")
30. 1or2or3or4or5or6or7or8or9	30. 1or2or3or4or5or6or7or8or9	29. 1or2or3or4or5or6or7or8	30. 1or2or3or4or5or6or7or8or9
31. 10or11or12or13or14or15or16	31. 10or11or12or13or14or15or16	30. 9or10or11or12or13or14or15	31. 10or11or12or13or14or15or16
32. 17or18or19or20or21or22or23or24or25	32. 17or18or19or20or21or22or23or24or25	31. 16or17or18or19or20or21or22or23or24	32. 17or18or19or20or21or22or23or24or25
33. 26or27or28or29	33. 26or27or28or29	32. 25or26or27or28	33. 26or27or28or29
34. 30 and 31 and 32 and 33	34. 30 and 31 and 32 and 33	33. 29 and 30 and 31 and 32	34. 30 and 31 and 32 and 33

Limt all to 1999-Current and English Language

Appendix 3: Scenario study scenarios

(All events and persons described in these scenarios are fictitious)

Scenario 1

Doris Campbell, an 80-year-old lady, was admitted to the acute medical unit with a urinary tract infection and acute confusion (delirium). Doris was assessed by the physiotherapist who noted reduced balance and the need for assistance to stand. The physiotherapist was going to pass this information onto the nurse but was asked to urgently provide chest physiotherapy to another patient who had become unwell. The nurse looking after Doris completed her falls risk assessment which suggested she was 'high risk'; among other things this prompted that Doris should have a falls sensor; the nurse had a lot to do and asked if the healthcare assistant could arrange for one to be fitted. The healthcare assistant didn't feel comfortable with how to fit a falls sensor and didn't want to trouble the busy nurse; they decided to find someone else to help when they got a moment.

No/low-harm outcome

Later that day Doris was found on the floor next to her bed. She was immediately assessed by the nurse and junior doctor. Although dazed Doris had fortunately not suffered any injuries and was helped back to her bed.

Severe harm outcome

Later that day Doris was found on the floor next to her bed. She was immediately seen by the nurse and junior doctor. Doris was dazed and complaining of pain in her left hip. An x-ray confirmed she had broken her leg and had to remain in hospital for this to be repaired. She was eventually discharged and made a good recovery.

Death outcome

Later that day Doris was found on the floor next to her bed. She was immediately seen by the nurse and junior doctor. Doris was initially agitated but rapidly became drowsy. She underwent a CT scan of her head, which showed a large bleed around her brain. Doris unfortunately died the following day with her family by her side.

Scenario 1

Following an investigation, the following factors were thought to have contributed to this incident:

- Doris was very confused and agitated putting her at greater risk of a fall
- Doris was placed in a bed without a clear line of site to the staff on the ward
- There were insufficient nursing staff on duty the day Doris was admitted
- The physiotherapist who noted the patient's falls risk had been distracted and so had not shared this with the nurse.
- The nurse asked the healthcare assistant to fit the falls sensor without realising they didn't feel confident in doing this.
- Doris was on a number of medications before she was admitted, which could increase her falls risk. The junior doctor who initially saw Doris was not sure which of these medications to stop and so decided to leave this for the consultant review.
- The brakes on Doris' bed had not been applied, which moved as she had tried to stand.
- There is no protocol for the management of confused patients.

Recommendations

- The healthcare assistant should undergo a period of supervised practice. [PUNITIVE]
- Consider disciplinary process for the nurse in charge of this patients care.[PUNITIVE]
- Remind all staff that brakes should be re-applied after moving a bed. [WEAK]
- Junior doctors to receive training on medications which increase fall's risk.[WEAK]
- Introduce a checklist to ensure that confused patients receive appropriate investigations and management, including those to reduce falls risk. [MEDIUM]
- Introduce an electronic alert within the computer based prescribing system to identify those medications, which increase fall's risk. [MEDIUM]
- New beds will be sourced which automatically apply brakes, which then have to be dis-engaged to move the bed. [STRONG]
- Continue the work of the Falls Reduction Committee to continuously improve falls reduction strategies such as low beds, zipper quilts, and falls action plans. [STRONG]
- None. I do not wish to select any recommendations

Scenario 2

David Carter, a 72 year old man, was admitted to the acute medical unit at his local hospital. Over the last few days he had been getting some pain in his abdomen but was otherwise well. He was seen by the junior doctor (Dr A) who ordered some blood tests and an x-ray of his abdomen and chest. David's observations were normal and he looked quite well. After an hour the consultant on duty reviewed David and Dr A's plan; they agreed with the junior doctor and wrote that the x-rays would help check for any evidence of perforation in David's bowels. The consultant did not directly discuss the case with Dr A, but suggested in the written notes that a surgical review should be arranged. The x-rays were not scheduled as urgent but would be done within a few hours. Shortly before the end of Dr A's shift there was an emergency in an adjacent ward, which they assisted with. Following this emergency Dr A went to hand over, during which the doctors hand over patients and jobs to the next shift of doctors. David Carter's x-rays had not yet been completed and Dr A failed to handover these for the incoming team to chase.

No/low-harm outcome

David remained well over night. When the consultant came to review David in the morning they noted that the x-rays showed that David did have a perforation in his bowel. After discussion with the surgical team they decided that David did not require surgery for this but transferred him to their care. He remained in hospital for a further day before being discharged. No harm resulted from the delay in x-rays being done or reviewed. The consultant spoke with Dr A, who admitted that they had forgotten to hand over the task of chasing David's x-rays.

Severe harm outcome

Later that night David became unwell and the nursing staff asked the doctors to review him. The doctors were very concerned about David and they noticed that he had had x-rays showing a perforation in his bowel. They immediately spoke to the surgical team, and he underwent emergency surgery. The delay in diagnosis of perforation made the operation more complicated and David remained in hospital for a further week. He was discharged home and has now made a full recovery. The consultant spoke with the Dr A, who admitted that they had forgotten to hand over the task of chasing David's x-rays.

Death outcome

Later that night David became unwell and the nursing staff asked the doctors to see him. The doctors were very concerned about David and they noticed that he had had x-rays showing a perforation in his bowel. They immediately spoke to the surgical team. David deteriorated very quickly and suffered a cardiac arrest (his heart stopped). Despite the efforts of the medical team David could

not be resuscitated and passed away. The consultant spoke with the Dr A, who admitted that they had forgotten to hand over the task of chasing David's x-rays.

Scenario 2

Following an investigation, the following factors were thought to have contributed to this incident:

- David was admitted to the acute medical unit but should have been admitted to the surgical assessment unit where staff are more familiar with perforation of the bowel.
- David was not displaying the typical physiological signs of perforation such as low blood pressure and tachycardia, and his abdomen was soft and only mildly tender which is again, not typical of perforation.
- Dr A was less familiar with surgical conditions.
- The consultant did not directly discuss David's case with Dr A or handover their concerns about perforation and the need to discuss the case with the surgical team.
- Dr A faced multiple distractions during their shift and in particular the emergency that occurred just before handover, when Dr A reports they would usually write a list of items to handover.
- There is no system of alerting staff when out-of-hours x-rays have been completed.
- When x-rays are completed with obvious abnormalities there is no system of alerting the parent medical team.
- There is no structure or standardised approach to the handover process.

Recommendations

1. Dr A should be suspended and further investigation into their practice carried out. [PUNITIVE]
2. Dr A will complete training on how to organise and prioritise tasks, including those for handover. [PUNITIVE]
3. Medical staff will receive training on surgical conditions, including perforation. [WEAK]
4. A new policy for out-of-hours x-rays will be produced. [WEAK]
5. The handover process will be standardised and a checklist developed to ensure all important information is discussed, including a prompt to handover outstanding x-rays. [MEDIUM]
6. Radiographers to bleep requesting team when out-of-hours x-rays are completed to prompt review. [MEDIUM]
7. All patients with abdominal pain will be admitted under the surgical team for initial assessment. [STRONG]

8. Hire an advanced radiographer for out-of-hours to ensure that all x-rays are reviewed and abnormalities are discussed with parent medical team. [STRONG]
9. None. I do not wish to select any recommendations

Scenario 3

Samantha Giles, a 32-year-old teacher, attended her local emergency department feeling short of breath. Doctor B who assessed Samantha thought she had a blood clot in her lungs (pulmonary embolism or 'PE'). The doctor prescribed a blood thinning injection (Dalteparin); the dose of which is calculated based on weight. There are different dose syringes based on the weight of the patient. After writing the prescription the doctor asked one of the nurses to administer the injection while they arranged a scan to confirm whether or not there was a PE. Shortly after giving Samantha the dalteparin injection, the nurse realised they had picked up the wrong dose syringe and had in fact given her a much higher dose than intended. The nurse and doctor explained that Samantha had had a much higher dose than intended and that there was a risk of bleeding. Samantha was kept in hospital and monitored closely.

No/low-harm outcome

Samantha was kept in hospital and monitored closely. No bleeding or harm occurred, and Samantha was discharged the following day.

Severe harm outcome

Samantha suddenly became unwell and vomited a large amount of blood. She had an urgent camera test to look into her stomach (endoscopy) which showed a bleeding ulcer. She lost a large amount of blood and had to remain in hospital for two more days than she would have done, had the bleed not happened.

Death outcome

Samantha suddenly became confused and had a seizure. She became unconscious and was sent for an urgent scan of her head. The scan showed that Samantha had a large bleed inside her head. She was urgently transferred to a neurosurgery centre for treatment but unfortunately died from complications of the surgery.

Scenario 3

Following an investigation, the following factors were thought to have contributed to this incident:

- This incident was investigated and it was felt the following factors contributed to the wrong dose:
- The emergency department did not have its full quota of nursing staff on duty, and so the nurse involved in this incident was stressed, rushed and distracted.
- The emergency department was particularly busy with higher than usual numbers of patients.
- The dalteparin injection comes in a variety of different dose syringes, but they all look very similar and are kept together
- The blood thinning injection syringes are kept in a cupboard under poor lighting
- The national 4-hour emergency department target places pressure on staff to work more quickly

Recommendations

1. The nurse involved should be suspended and further investigation into their practice carried out. [PUNITIVE]
2. The nurse should undergo a period of supervised practice. [PUNITIVE]
3. Training for emergency department staff on the risks of blood thinning medications. [WEAK]
4. Place a warning poster on the cupboard containing the blood thinning medications, to remind staff to select the correct strength syringe. [WEAK]
5. Modify the labelling on the different dose strength blood thinning medication syringes so that they look significantly different. [MEDIUM]
6. Provide emergency department staff with regular opportunities to carry out tasks under pressure in a simulated environment, with subsequent feedback/debrief. [MEDIUM]
7. Switch to a different supplier of blood thinning medication syringes that have been designed specifically to reduce the risk of selecting the wrong syringe. [STRONG]
8. Senior leadership to set up a working group and commit to a range of projects to improve patient safety in the emergency department. [STRONG]
9. None. I do not wish to select any recommendations

Appendix 4: Scenario study example questionnaire

Each online questionnaire contained three separate scenarios – only one scenarios is presented here to show the structure of the online questionnaire.

Scenario Study - GAR (V 10-2-22)

What's the purpose of this research?

It is estimated that 1 in 10 patients are harmed as the result of healthcare[1]; and that this is often the result of both human error and poor system design. One of the ways in which safety might be improved is to investigate these patient safety incidents (PSIs). PSIs are routinely investigated within healthcare, with the purpose of finding why the error or harm occurred. Following an investigation staff will usually identify some actions or changes that could be introduced to reduce the chance the incident is repeated or to improve safety; so called recommendations. We are interested in looking at what kind of recommendations people select.

We will present you with three fictional incidents involving a fall, delay in diagnosis and medication error. We will then ask you some questions about how we should respond to the incident and to consider some possible changes we could make to prevent it happening again. For the purposes of this study we would like you to ignore potential costs, or difficulty in putting into practice; imaging you have infinite funds and resources.

You will need to:
1) read about three incidents
2) answer some questions about how we should respond

1 Braithwaite J (2018) Changing how we think about healthcare improvement. BMJ 2018;361:k2014

Consent

Before we start we would like to check that you are happy to take part in this research study.

You are being invited to participate in a research study looking at what recommendations people select following medical errors/adverse events. This study is being done by William Lea*[1], Jane O'Hara*[1], Rebecca Lawton*[1] and Charles Vincent*[2].

We believe there are no known risks associated with this research study; however, as with any online related activity the risk of a breach is always possible. To the best of our ability your participation in this study will remain confidential, and only anonymised data will be published.

- I understand that my participation is voluntary and that I am free to withdraw at any time prior to submitting the completed questionnaire without giving any reason and without there being any negative consequences.
- I understand that my participation or my individual responses in this study will not be shared with my employer.
- I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research.
- I agree for the data collected from me to be used in relevant future research in an anonymised form.

*[1] University of Leeds, *[2] University of Oxford

I agree with the above statements and I am happy to take part in this research:

- ○ Yes
- ○ No

Further information about your rights in relation to the information we collect can be found here: https://dataprotection.leeds.ac.uk/wp-content/uploads/sites/48/2020/11/Research-Participant-Privacy-Notice.docx

A little bit about you

How old are you?

Would you describe yourself as a

- Women
- Man
- Non-binary
- other

Prefer to self describe as

Do you work in healthcare?

- Yes
- No

A little bit about you (Non-Healthcare)

Have you ever been involved in a patient safety incident/adverse event as a:

	Yes	No	approximately how many times?	comments
Victim (patient or family member)				
None of the ways listed above (please give details in comments box)				

Have you ever had training or do you have expertise in patient safety or how to investigate adverse events? (e.g. carry out investigations, undertake relevant research)

- Yes
- No

If you have answered 'yes' please could you provide some details, such as job role/research area or content and length of training.

A little bit about you (Healthcare)

Have you ever been involved in a patient safety incident/adverse event as a:

	Yes	No	approximately how many times?	comments
Victim (patient or family member)	○	○		
Member of staff involved in the incident	○	○		
Member of the investigation team	○	○		
None of the ways listed above (please give details in comments box)	○	○		

Have you ever had training or do you have expertise in patient safety or how to investigate adverse events? (e.g. carry out investigations, undertake relevant research)

- ○ Yes
- ○ No

If you have answered 'yes' please could you provide some details, such as job role/research area or content and length of training.

What is your profession?

If you selected Other, please specify:

What is your grade? (e.g. Band 3, Foundation Year 1 doctor)

What year did you qualify?

Scenario 1

A fall on the ward: a brief summary

Doris Campbell, an 80 year old lady, was admitted to the acute medical unit with a urinary tract infection and acute confusion (delirium). Doris was assessed by the physiotherapist who noted reduced balance and the need for assistance to stand. The physiotherapist was going to pass this information onto the nurse but was asked to urgently provide chest physiotherapy to another patient who had become unwell. The nurse looking after Doris completed her falls risk assessment which suggested she was 'high risk'; among other things this prompted that Doris should have a falls sensor(i); the nurse had a lot to do and asked if the healthcare assistant could arrange for one to be fitted. The healthcare assistant didn't feel comfortable with how to fit a falls sensor and didn't want to trouble the busy nurse; they decided to find someone else to help when they got a moment.

Later that day Doris was found on the floor next to her bed. She was immediately assessed by the nurse and junior doctor. Although dazed Doris had fortunately not suffered any injuries and was helped back to her bed.

Following an investigation, the following factors were thought to have contributed to this incident:

1. Doris was very confused and agitated putting her at greater risk of a fall
2. Doris was placed in a bed without a clear line of site to the staff on the ward
3. There were insufficient nursing staff on duty the day Doris was admitted
4. The physiotherapist who noted the patient's falls risk had been distracted and so had not shared this with the nurse.
5. The nurse asked the healthcare assistant to fit the falls sensor without realising they didn't feel confident in doing this.
6. Doris was on a number of medications before she was admitted, which could increase her falls risk. The junior doctor who initially saw Doris was not sure which of these medications to stop and so decided to leave this for the consultant review.
7. The brakes on Doris' bed had not been applied, which moved as she had tried to stand.
8. There is no protocol for the management of confused patients.

Q1 How responsible do you feel the healthcare staff are in causing this incident?

Please don't select more than 1 answer(s) per row.

	1	2	3	4	5
1 = not at all , 5 = entirely	☐	☐	☐	☐	☐

Q2 How avoidable do you think this incident was?

Please don't select more than 1 answer(s) per row.

	1	2	3	4	5
1 = not at all, 5 = entirely	☐	☐	☐	☐	☐

Q3 How important do you think it is for an investigation to be carried out?

Please don't select more than 1 answer(s) per row.

	1	2	3	4	5
1 = not at all, 5 =very	☐	☐	☐	☐	☐

Q4 What level of harm was caused by this incident?

- ○ No/Low harm
- ○ Severe harm
- ○ Death

Recommendations

Please consider the following recommendations and select **up to five** to reduce the chance this incident happens again.

Please select no more than 5 answer(s).

- ☐ The healthcare assistant should undergo a period of supervised practice.
- ☐ Consider disciplinary process for the nurse in charge of this patients care.
- ☐ Remind all staff that brakes should be re-applied after moving a bed.
- ☐ Junior doctors to receive training on medications which increase fall's risk.
- ☐ Introduce a checklist to ensure that confused patients receive appropriate investigations and management, including those to reduce falls risk.
- ☐ Introduce an electronic alert within the computer based prescribing system to identify those medications, which increase fall's risk.
- ☐ New beds will be sourced which automatically apply brakes, which then have to be dis-engaged to move the bed.
- ☐ Continue the work of the Falls Reduction Committee to continuously improve falls reduction strategies such as low beds, zipper quilts, and falls action plans.
- ☐ None. I do not wish to select any recommendations

[Optional] Please include any explanation for your choices here or suggest other recommendations you might have proposed:

TWO MORE SCENARIOS FOLLOWED IN ONLINE QUESTIONNAIRE

Did you face any difficulties answering these questions?

Did you find any of the questions difficult to answer? Or do you have any other comments you would like to make?

Thank you

We are very grateful for your time in taking part in this research, and hope that the findings will improve the way in which we can learn from errors within healthcare.

If you have any questions about the research or would like to be informed about the results please contact William Lea (umwl@leeds.ac.uk)

Taking part in this survey may have raised questions for you about incidents and harm. The following organisations provide information and support:

Action Against Medical Accidents (www.avma.org.uk): whose purpose if to support people affected by avoidable harm in healthcare; to help them achieve hustice; and to promote better patient safety for all.

Harmed Patient Alliance (www.harmedpatientsalliance.org.uk): are working to provide support for patients and families following harm.

Second Victim Support (www.secondvictim.co.uk): recognises the impact that a patient safety incident has on the healthcare employee involved. Whilst the patient and their family are always the priority, this website seeks to help 'second victims' identify the types of support they may need, and signpost them towards help.

Appendix 5: Scenario study multilevel modelling results

Supplemental Table: Full Model Results

Outcome	Model type	Predictor	Contrast	Estimate [95% CI]	R^2_M | R^2_C	ICC	BIC
Number of recommendations	Model 1: adjusted for participant age + gender	Outcome	Severe vs. No/low-harm	-0.05 [-0.24, 0.13]	1.73% | 40.33%	39.28%	1680
			Death vs. No/low-harm	**0.20** [0.02, 0.39]**			
		Age	-	-0.01 [-0.02, 0.00]			
		Gender	Male vs. Female	-0.15 [-0.46, 0.17]			
	Model 2: Model 1 variables + adjusted for scenario (1=fall, 2=x-ray, 3=drug dose) + participant group (0=public, 1=staff, 2=experts)	Outcome	Severe vs. No/low-harm	-0.07 [-0.24, 0.11]	5.68% | 43.73%	40.35%	1689
			Death vs. No/low-harm	**0.20** [0.02, 0.38]**			
		Age	-	0.00 [-0.01, 0.01]			
		Gender	Male vs. Female	-0.16 [-0.48, 0.16]			
		Scenario	**X-ray vs. Fall**	**-0.28*** [-0.46,-0.10]**			
			Drug vs. Fall	**-0.41*** [-0.58,-0.23]**			
		Participant group	Staff vs. Public	0.10 [-0.20, 0.41]			
			Expert vs. Public	**-0.35*[-0.71, 0.00]**			
	Model 3: Model 2 variables + participant previous incident involvement (none=0,victim=1, staff member=2, both=3)	Outcome	Severe vs. No/low-harm	-0.07 [-0.24, 0.11]	7.38% | 44.24%	39.79%	1706
			Death vs. No/low-harm	**0.20** [0.02, 0.38]**			
		Age	-	0.00 [-0.01, 0.01]			
		Gender	Male vs. Female	-0.13 [-0.45, 0.19]			
		Scenario	**X-ray vs. Fall**	**-0.28*** [-0.46,-0.10]**			
			Drug vs. Fall	**-0.41*** [-0.58,-0.23]**			
		Participant group	Staff vs. Public	0.23 [-0.15,0.61]			
			Expert vs. Public	-0.26 [-0.68,0.16]			

Supplemental Table: Full Model Results

Outcome	Model type	Predictor	Contrast	Estimate [95% CI]	R^2_M | R^2_C	ICC	BIC
		Involvement	Victim vs. None	0.19 [-0.15,0.54]			
			Staff member vs. None	-0.05 [-0.45, 0.35]			
			Both vs. None	-0.50~ [-1.06, 0.05]			
Recommendation Score	Model 1: adjusted for participant age + gender				1.00% | 33.00%	28.56%	819
		Outcome	Severe vs. No/low-harm	-0.07 [-0.15, 0.02]			
			Death vs. No/low-harm	-0.08 [-0.17, 0.00]			
		Age	-	0.00 [-0.00, 0.00]			
		Gender	Male vs. Female	0.03 [-0.11, 0.17]			
	Model 2: Model 1 variables + adjusted for scenario (1=fall, 2=x-ray, 3=drug dose) + participant group (0=public, 1=staff, 2=experts)				12.0% | 39.0%	26.8%	807
		Outcome	Severe vs. No/low-harm	-0.06 [-0.15, 0.02]	P=0.057		
			Death vs. No/low-harm	-0.08 [-0.16, 0.00]			
		Age	-	0.00 [-0.00, 0.00]			
		Gender	Male vs. Female	0.01 [-0.12, 0.14]			
		Scenario	**X-ray vs. Fall**	**-0.25*** [-0.33, -0.17]**			
			Drug vs. Fall	-0.05 [-0.14, 0.03]			
		Participant	Staff vs. Public	0.07 [-0.05, 0.20]			
		group	**Expert vs. Public**	**0.37***[0.21, 0.53]**			
	Model 3: Model 2 variables + participant previous incident involvement (none=0,victim=1, staff member=2, both=3)				14.0% | 40.0%	26.8%	829
		Outcome	Severe vs. No/low-harm	-0.06 [-0.15, 0.02]			
			Death vs. No/low-harm	-0.08 [-0.16, 0.00]			
		Age	-	0.00 [-0.00, 0.00]			
		Gender	Male vs. Female	-0.00 [-0.13, 0.13]			
		Scenario	**X-ray vs. Fall**	**-0.25*** [-0.33, -0.17]**			
			Drug vs. Fall	-0.05 [-0.14, 0.03]			

Supplemental Table: Full Model Results

Outcome	Model type	Predictor	Contrast	Estimate [95% CI]	R^2_M | R^2_C	ICC	BIC
		Participant group	Staff vs. Public	-0.02 [-0.18, 0.13]			
			Expert vs. Public	**0.27*** [0.13, 0.40]**			
		Involvement	Victim vs. None	-0.08 [-0.22,0.05]			
			Staff member vs. None	0.14 [-0.04, 0.31]			
			Both vs. None	0.18 [-0.04, 0.41]			
Responsibility	Model 1: adjusted for participant age + gender	Outcome	**Severe vs. No/low-harm**	**0.17** [0.02, 0.32]**	2.55% | 52.13%	50.87%	1549
			Death vs. No/low-harm	**0.25*** [0.10, 0.41]**			
		Age	-	0.00 [-0.01, 0.01]			
		Gender	Male vs. Female	-0.30 [-0.62, 0.01]			
	Model 2: Model 1 variables + adjusted for scenario (1=fall, 2=x-ray, 3=drug dose) + participant group (0=public, 1=staff, 2=experts)	Outcome	**Severe vs. No/low-harm**	**0.17** [0.02, 0.32]**	12.46% | 53.67%	47.07%	1547
			Death vs. No/low-harm	**0.26*** [0.11, 0.41]**			
		Age	-	0.00 [-0.01, 0.01]			
		Gender	Male vs. Female	-0.24 [-0.54, 0.06]			
		Scenario	**X-ray vs. Fall**	**-0.26*** [-0.42, -0.11]**			
			Drug vs. Fall	-0.14 [-0.29, 0.01]			
		Participant group	Staff vs. Public	-0.28~ [-0.57, 0.01]			
			Expert vs. Public	**-0.90***[-1.23, -0.56]**			
	Model 3: Model 2 variables + participant previous incident involvement (none=0,victim=1, staff member=2, both=3)	Outcome	**Severe vs. No/low-harm**	**0.17** [0.02, 0.32]**	15.51% | 54.05%	45.62%	1560
			Death vs. No/low-harm	**0.26*** [0.11, 0.41]**			
		Age	-	0.00 [-0.01, 0.01]			
		Gender	Male vs. Female	-0.23 [-0.52, 0.06]			

Supplemental Table: Full Model Results

Outcome	Model type	Predictor	Contrast	Estimate [95% CI]	R^2_M | R^2_C	ICC	BIC
		Scenario	**X-ray vs. Fall**	**-0.26*** [-0.42, -0.11]**			
			Drug vs. Fall	-0.14~ [-0.29, 0.01]			
		Participant	**Staff vs. Public**	**-0.38** [-0.72, -0.03]**			
		group	**Expert vs. Public**	**-1.00*** [-1.39, -0.62]**			
		Involvement	**Victim vs. None**	**0.52*** [0.20, 0.83]**			
			Staff member vs. None	0.20 [-0.17, 0.57]			
			Both vs. None	0.33 [-0.18, 0.85]			
Avoidability	Model 1: adjusted for participant age + gender	Outcome	Severe vs. No/low-harm	-0.05 [-0.22, 0.11]	0.93% | 32.35%	31.71%	1499
			Death vs. No/low-harm	-0.04 [-0.21, 0.12]			
		Age	-	0.00 [0.00, 0.01]			
		Gender	Male vs. Female	-0.16 [-0.41, 0.09]			
	Model 2: Model 1 variables + adjusted for scenario (1=fall, 2=x-ray, 3=drug dose) + participant group (0=public, 1=staff, 2=experts)	Outcome	Severe vs. No/low-harm	-0.04 [-0.19, 0.11]	10.46% | 39.70%	32.66%	1477
			Death vs. No/low-harm	-0.03 [-0.19, 0.12]			
		Age	-	0.00 [-0.01, 0.01]			
		Gender	Male vs. Female	-0.09 [-0.33, 0.16]			
		Scenario	X-ray vs. Fall	0.04 [-0.11, 0.20]			
			Drug vs. Fall	**0.47*** [0.31, 0.62]**			
		Participant	**Staff vs. Public**	**-0.37*** [-0.61, -0.13]**			
		group	**Expert vs. Public**	**-0.49***[-0.76, -0.22]**			
	Model 3: Model 2 variables + participant previous incident involvement	Outcome	Severe vs. No/low-harm	-0.04 [-0.19, 0.11]	11.24% | 40.21%	32.64	1498
			Death vs. No/low-harm	-0.03 [-0.19, 0.12]			
		Age	-	0.00 [-0.01, 0.01]			

Supplemental Table: Full Model Results

Outcome	Model type	Predictor	Contrast	Estimate [95% CI]	R^2_M | R^2_C	ICC	BIC
	(none=0,victim=1, staff member=2, both=3)	Gender	Male vs. Female	-0.07 [-0.31, 0.18]			
		Scenario	X-ray vs. Fall	0.04 [-0.11, 0.20]			
			Drug vs. Fall	**0.47*** [0.31, 0.62]**			
		Participant	Staff vs. Public	-0.25~ [-0.55, 0.04]			
		group	**Expert vs. Public**	**-0.38** [-0.70, -0.05]**			
		Involvement	Victim vs. None	0.11 [-0.15, 0.38]			
			Staff member vs. None	-0.19 [-0.50, 0.12]			
			Both vs. None	-0.15 [-0.58, 0.28]			
Importance of Investigating	Model 1: adjusted for participant age + gender	Outcome	**Severe vs. No/low-harm**	**0.38*** [0.24, 0.51]**	10.46% | 47.89%	41.81%	1345
			Death vs. No/low-harm	**0.63*** [0.50, 0.76]**			
		Age	**-**	**0.01** [0.00, 0.02]**			
		Gender	Male vs. Female	0.06 [-0.18, 0.30]			
	Model 2: Model 1 variables + adjusted for scenario (1=fall, 2=x-ray, 3=drug dose) + participant group (0=public, 1=staff, 2=experts)	Outcome	**Severe vs. No/low-harm**	**0.39*** [0.26, 0.51]**	13.30% | 51.93%	44.55%	1353
			Death vs. No/low-harm	**0.64*** [0.51, 0.77]**			
		Age	**-**	**0.01** [0.00, 0.02]**			
		Gender	Male vs. Female	0.04 [-0.20, 0.28]			
		Scenario	X-ray vs. Fall	-0.11 [-0.23, 0.02]			
			Drug vs. Fall	**0.24*** [0.11, 0.37]**			
		Participant	Staff vs. Public	0.11 [-0.13, 0.34]			
		group	Expert vs. Public	-0.05[-0.32, 0.22]			
	Model 3: Model 2 variables + participant previous incident	Outcome	**Severe vs. No/low-harm**	**0.38*** [0.26, 0.51]**	14.97% | 52.35%	43.96%	1371
			Death vs. No/low-harm	**0.64*** [0.51, 0.77]**			

Supplemental Table: Full Model Results

Outcome	Model type	Predictor	Contrast	Estimate [95% CI]	R^2_M | R^2_C	ICC	BIC
	involvement (none=0,victim=1, staff member=2, both=3)	Age	-	**0.01** [0.00, 0.02]**			
		Gender	Male vs. Female	0.06 [-0.18, 0.30]			
		Scenario	X-ray vs. Fall	-0.11 [-0.23, 0.02]			
			Drug vs. Fall	**0.24*** [0.11, 0.37]**			
		Participant	Staff vs. Public	0.11 [-0.18, 0.40]			
		group	Expert vs. Public	-0.07 [-0.38, 0.25]			
		Involvement	**Victim vs. None**	**0.31** [0.04, 0.57]**			
			Staff member vs. None	0.08 [-0.23, 0.39]			
			Both vs. None	-0.06 [-0.48, 0.36]			