BEYOND EBM

A THESIS ON THE FUNDAMENTALS OF CLINICAL KNOWLEDGE CREATION

SIETSE WIERINGA, GP MSC.



Thesis submitted for the degree of
Doctor of Philosophy in Evidence Based Health Care

Kellogg College, University of Oxford
2019

For Tara

Thesis Abstract

Introduction

The evidence-based medicine (EBM) movement has claimed to have induced and improved the production of trustworthy clinical guidelines. However, ethnographic research showed that clinicians rarely use explicit evidence from guidelines directly. Instead they draw heavily on socially shared knowledge and patterns of behaviour, called 'mindlines'. To understand how EBM reaches its limits in producing meaningful guidance and finding what may lie beyond, this thesis explores the tensions between guidelines and mindlines from a philosophy of science perspective.

Methods

The approach I take to uncover these tensions is to view them as anomalies of a mature Kuhnian paradigm. Unlike most previous scholars, I chose not to study the use of guidelines in clinical practice but the guideline development process as a central means of the EBM paradigm to create decision making support for individual patients instead. The project draws on theoretical findings from a literature review of mindlines and philosophical analyses of perceptions in EBM about knowledge, reasoning and truth, which are juxtaposed and synthesised with empirical data from a digital ethnography of mindlines in virtual social networks of clinicians and an ethnography of guideline panels in the United Kingdom, Norway and the Netherlands.

Results

I show how mindlines have been researched and acknowledged in multiple contexts and

evolved conceptually as a container for many theories of knowledge. In a critique I consider how mindlines challenge the fundamentals of EBM. Taking a Latour inspired perspective, I elicit how a changing EBM is trying to cope with the downsides of what he calls “purification”: a scientific practice of categorising and quantifying reality. I then provide insights into the challenges of EBM for guideline developers concerning ways to make inferences and integrate evidence. It highlights important issues in guideline development regarding its purpose, use and integration of different knowledges and suggests alternatives for the dominant reasoning with RCTs. Next, drawing on work by Isabelle Stengers, I consider bias and the ideal limit theorem as a theory of truth in EBM. I argue that debates about truth remain hidden behind debates on bias and contend that the latter is in a dual, complex, necessary, unproductive as well as a productive conjunction with truth. I present characteristics of the ideal limit theorem as a dominant theory of truth in EBM and look at several other theories of truth that operate in the clinical encounter.

I then report the ethnographic findings on how all these issues played out in some real examples of informal knowledge creation and sharing in large groups of clinicians, and in formal guideline production panels. I provide an analysis of the characteristics of informal knowledge in virtual social networks of clinicians in the United Kingdom, Norway and the Netherlands, which sees it as produced collectively by clinicians to base their decisions on; a concept for explicit as well as tacit knowing shared among social groups and reinforcing norms of good practice in a fluid, dynamic and constantly evolving way.

Analysis of the findings of guideline panels in the UK, Norway and the Netherlands indicate that there are several tensions regarding mindlines and guidelines, most notably originating from two linked yet contrasting epistemologies within the guideline community: a purist understanding of EBM, adhering to frequency type reasoning, the ideal limit theorem and strict explicit purification, and a more pluralist understanding of EBM embracing multiple kinds of knowledge, reasoning and truth.

Conclusion

My empirical and theoretical work was aimed at providing new insights in mindlines and the production of guidelines using an epistemological perspective on EBM as a mature and changing paradigm, with a focus on knowledge, reasoning, truth and purification. The implications for guideline production, teaching EBM and future research include a call to rethink the nature of knowledge and knowledge translation, to develop and examine new guideline tools including the use of virtual social networks, to discuss and critically appraise hidden assumptions regarding key concepts such as bias and pluralism, as well as to further theorise inductive inference in individual patients with the help of the philosophy of science as a proper academic field in healthcare.

Table of Contents

Thesis Abstract **2**

Acknowledgements **7**

Abbreviations **9**

Chapter 1: Beyond EBM **10**
1.1 Introduction 10
1.2 Aim and objectives 12
1.3 Research question 13
1.4 Personal history and motivation 14
1.5 Structure of this thesis 17

Chapter 2 Scope and philosophical approach **23**
2.1 Introduction 23
2.2 EBM as a paradigm 23
2.3 Shifting paradigms 30
2.4 The epistemology of EBM 34
2.5 Developing a framework for analysis 48
2.6 Conclusion 51

Chapter 3: Methodology and methods **52**
3.1 Methodology 52
3.2 Description of methods: meta-narrative literature review on mindlines 58
3.3 Description of methods: theoretical work on purification, inference and the concept of truth 63
3.4 Description of methods: a digital ethnography of mindlines online 66
3.5 Description of methods: ethnographic study of guideline development panels 78
3.6 Conclusion 88

Chapter 4: On the concept of knowledge and knowledge creation: mindlines **90**
4.1 Introduction 90
4.2 Results 90
4.3 Discussion 100
4.4 Conclusion 111

Chapter 5: On purification: has evidence-based medicine ever been modern? **113**
5.1 Introduction 113
5.2 'We have never been modern' 114
5.3 Early EBM as a modernist movement 117
5.4 The 'movement in crisis': EBM's receding modernist vision 120
5.5 EBHC – moving towards a 'middle kingdom'? 122
5.6 Discussion 126
5.7 Conclusion 132

Chapter 6: On inference: evading the problem of induction **133**
6.1 Introduction: The challenge of knowledge inclusion in guidelines 133
6.2 The purpose of guideline development 135
6.3 The problem of induction 137
6.3 The dominance of frequency-based reasoning 138
6.4 The challenge of integration 142
6.5 Conclusion 145

Chapter 7: Rethinking bias and truth **148**
7.1 Introduction 148
7.2 The ideal limit theorem of truth 150
7.3 Bias as inter-esse 153
7.4 Duality of bias challenging the ideal limit 156
7.5 Towards a pluralistic understanding of truth in EBM 160
7.6 Analysis 164
7.8 Conclusion 168

Chapter 8 Mindlines online: the nature of informal clinical knowledge **169**
8.1 Introduction 169
8.2 Results 170
8.3 Discussion 184
8.4 Conclusions 189

Chapter 9: An ethnographic study of guideline development panels in three countries **190**
9.1 Introduction 190
9.2 Results 190
9.3 Discussion 205
9.4 Conclusion 208
Examples from guideline development processes in the UK, the Netherlands and Norway 209

Chapter 10: Discussion and conclusions **218**
10.1 Introduction 218
10.2 Summary of key findings 218
10.3 Reflections and limitations of methodology 222
10.4 Implications and further research 227
10.5 Final remarks 237

Literature **239**

Acknowledgements

Throughout this thesis I have received much assistance and support from so many. Foremost I thank my supervisors Trisha Greenhalgh in Oxford and Eivind Engebretsen and Kristin Heggen in Oslo, whose considerations, critical remarks, and inspiration have been invaluable. Thank you for giving me this opportunity.

I thank John Gabbay and Andrée le May for their help and insights on mindlines that formed a stepping stone for this thesis. I acknowledge and thank Dunja Dreesens, Frode Forland, Carel Hulshof, Sue Lukersmith, Fergus Macbeth, Beth Shaw, Arlène van Vliet and Teun Zuiderent-Jerak on behalf of the AID Knowledge Working Group of the Guidelines International Network for their collaboration. Teun and Beth were instrumental in helping me to understand guideline production and getting me in contact with key persons. Thank you both so much. For the work presented in Chapter 8 I thank all the members of the online networks, as well as their administrators Kartik Modha from TGG, Hogne Sandvik from Eyr and especially Guido Verboom at NHG. Kristiane M. Hansson, David Wright and Gina Fraas Henrichsen for their invaluable contributions towards data collection. I thank Farida Vis and Natasha Whiteman of Manchester University for their help and advice regarding ethical issues in digital ethnography. The KNOWIT group in Oslo for their helpful input. A special thank you for Toni Tan and Nichole Taske at NICE, Margriet Bouma at NHG and Frode Forland in Oslo for making this research possible.

I wish to thank all my dear friends and colleagues who stood by me. Tamara Platteel for the countless nights we discussed philosophy and the meaning of life, Paulus Lips for your thoughts and introducing me to Kuhn, Latour and social constructivism. I thank Dee Hora, Alexis Cambanis, Dan Bernstein and Rob Meijer for their intelligence, deliberations and most importantly, for their friendship. I thank Claudia van Dijk and the other amazing and thoughtful GPs from Generation Next who work so hard to make primary care even better. Babette Rump for getting me to write up some of the early ideas on inference in healthcare, I hope you're doing well. Many thanks to Luisa Pettigrew, Stephanie Kumpunen, Greg Irving, Nick Fahy, Dominic Hurst and Sir Muir Gray who made British academia feel welcoming.

And finally, my family, my parents Joke and Wim, brother Folkert and sister Jantien for being always near, even when far away. Ervina, who was there for me throughout these often-chaotic years. I realise I often have not expressed my gratitude to you enough. And thank you Tara for teaching me that life is actually about watching ducklings in the Vondel Park, taking baby steps on the lawn and falling asleep while holding someone you love.

I acknowledge the National Institute for Health Research for funding my In-Practice Fellowship at Queen Mary University and the European Union Seventh Framework Programme for my Scientia scholarship at the University of Oslo (FP7-PEOPLE-2013-COFUND, Grant/Award Number: 609020).

Abbreviations

AID	Appraising and including different knowledge working group from GIN
CAT	Critically appraised topic
CERQual	Confidence in the Evidence from Reviews of Qualitative Research
DM	Diabetes Mellitus
DW	David Wright senior researcher, University of Southampton
EBHC	Evidence Based Healthcare
EBM	Evidence Based Medicine
EE	Professor Eivind Engebretsen, University of Oslo
FMLM	Faculty of Medical Leadership and Management
GFH	Gina Fraas Henrichesen, senior researcher at University of Oslo
GIN	Guidelines International Network
GOBSAT	Good Old Boys Sat Around a Table
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HAWeb	Virtual network set up by the professional GP societies in the Netherlands
HMIC	Health Management Information Consortium
IDREC	Social Sciences and Humanities Inter-divisional Research Ethics Committee at the University of Oxford
KH	Professor Kristin Heggen, University of Oslo
KMH	Kristiane M. Hansson, research fellow at the University of Oslo
KNOWIT	Knowledge in Translation research group at the University of Oslo
LDL	Low-Density Lipoproteins
LETR	Leading evidence to recommendations
LHV	Dutch National GP Association
NAFLD	Non-alcoholic fatty liver disease
NCBI	National Center for Biotechnology Information
NICE	National Institute for Health and Care Excellence
NHG	Dutch College of General Practitioners
NL	The Netherlands
PICO	Model for clinical questions: People, Intervention, Comparison, Outcome
QMREC	Queen Mary University Research Ethics Committee
RCT	Randomised Controlled Trial
REC	Research Ethics Committee
TFT	Thyroid function tests
TG	Professor Trisha Greenhalgh, University of Oxford
TGG	Tiko's GP Group, a group for general practitioners on Facebook
UK	The United Kingdom
WMO	Medical Research Involving Human Subjects Act in the Netherlands

Chapter 1: Beyond EBM

1.1 Introduction

The evidence-based movement is over 25 years old and has profoundly affected the way healthcare is practised. Originally known as evidence-based medicine (EBM) and now usually referred to as evidence-based healthcare (EBHC) to embrace all healthcare professions*, the movement has achieved much but also drawn criticism and concerns.

In 2014 a group of scholars questioned whether the EBHC movement was 'in crisis' as a result of a number of unintended consequences of its earlier success [1]:

> *"the evidence-based 'quality mark' has been misappropriated by vested interests; the volume of evidence, especially clinical guidelines, has become unmanageable; statistically significant benefits may be marginal in clinical practice; inflexible rules and technology driven prompts may produce care that is management driven rather than patient centred; and evidence-based guidelines often map poorly to complex multimorbidity".*

Whilst many evidence-based tools for generating, validating and disseminating medical knowledge have been developed and implemented to appraise evidence from randomised controlled trials (RCTs), we are as yet unable to fully assess and incorporate the range of evidence that is relevant to all the problems that clinicians and patients have to cope with in everyday practice. Knowledge from (for instance) "outbreak investigations, laboratory research, mathematical modelling, qualitative research, or quality improvement processes and clinical audit" are underrepresented in clinical guidance [2].

This was clearly not intended by the pioneers of EBM. They argued that clinical expertise and patient preferences should be integrated with the best evidence [3]. Moreover, the best evidence should not be limited by a simple and restrictive hierarchy of evidence, as some clinical issues are not best addressed using RCTs, some require guidance at a point in time and are not able to wait for RCTs to be undertaken, and some need a range of evidence to fully support the resulting recommendations. Nonetheless, methods of quality assessment and appraisal used in current guideline development, such as Grading of Recommendations Assessment, Development and Evaluation (GRADE), can drive the process towards appraising evidence from RCTs as the primary source of evidence, such that integrating knowledge from other sources is largely neglected.

As a result, critics have complained about the role of EBM in guideline development processes, such as a reductionist scope among other epistemological problems [4,5], limited use of heuristics [6], a lack of focus on the individual patient and shared decision

making, misuse of EBM for vested interests driving over-diagnosis and inability to deal with multi-morbidity [1,7].

Movements of scientists have recently emerged who seek to reform EBM, healthcare research and guideline development [2,8,9] to deal with these challenges. For example, the Guidelines International Network (GIN) has set up the Appraising and Including Different Kinds of Knowledge (AID) working group (http://www.g-i-n.net/working-groups/aid-knowledge), whose members include staff from National Institute for Health and Care Excellence (NICE), NHG (Dutch College of General Practitioners broadly equivalent to NICE in the Netherlands) and comparable organisations, in order to identify "methods and promising initiatives for appraising and including a wider range of knowledge sources in guidelines".

1.2 Aim and objectives

This thesis explores the concept and fundamentals of EBM as a practice of creating clinical knowledge. The concept of 'mindlines' has a central position in this thesis. This alternative model of knowledge creation resulted from ethnographic research by Gabbay and Le May [10] who found that clinicians rarely used explicit evidence from guidelines directly. Instead they drew heavily on socially shared knowledge and patterns of behaviour, called 'mindlines'. Influencing mindlines through guidelines, and vice versa, has proved to be challenging [11]. Exploring the idea of mindlines theoretically and empirically in guideline development organisations and the wider medical

community (using conversations exchanged on virtual social networks), may offer a glimpse of what lies beyond EBM.

1.3 Research question

In the context of growing awareness that knowledge from research findings, distilled into traditional clinical guidelines, is only one of many contributions to clinical decision-making, the research question addressed in this thesis is: How does the philosophy of knowledge help us to explore the tensions between guidelines and mindlines?

The scope of this thesis is confined to understanding how EBM reaches its limits in producing meaningful and valid guidance for clinical practice and finding what may lie beyond formal guideline development by professional institutions based on the principles of mindlines, epistemology and empirical evidence. I provide an understanding of the changing practice of EBM, the current concept of knowledge, ways to infer and perceptions of validity and truth. The aim of is to explore and expand the means of knowledge creation, including the use of virtual technology, thereby helping to ensure that guidelines interact in a meaningful way with the mindlines that exist as informal knowledge in the community of clinicians and patients and support evidence-based decision making in every day practice.

Throughout the thesis I will address the following sub-questions to reflect this scope:

- Whether and how have mindlines influenced, or should influence, the EBM movement?

- Does the emergence of EBM/EBHC movement mark a shift from the traditional 'priestly authority' of doctors to a more modern, knowledge-based – and hence rational and scientific – authority?
- What are the issues of using different knowledge and reasoning in guideline development?
- What are the fundamental beliefs and assumptions on bias and truth that drive and constrain what we can know and do in clinical practice?
- How is medical knowledge developed and shared in informal communities of clinicians?
- How do guideline development panels combine (or why do they find that they are unable to combine) different kinds of evidence?
- How is the work of guideline development panels influenced by different perspectives on epistemic hierarchies and approaches to uncertainty?

1.4 Personal history and motivation

The context that led me to the start of this thesis and applying for transfer to DPhil status has been gradually evolving from the early stages of my career. I studied medicine and graduated Cum Laude at Leiden University in the Netherlands. After working a couple of years in a hospital in Curacao I decided to specialise in general practice as I feel that this is where the very nature of healthcare is most apparent. To satisfy an urge to think about systems and broaden my scope of the concept of healthcare I did a master's in healthcare management at the Erasmus University in Rotterdam.

My main professional and academic interest lies in the creation of medical knowledge. How do we get to useable truths in medicine? What are they? And why? I first noticed the problematic nature of knowledge development when trying to improve diabetes and out of hours care in Curacao. In Curacao three medical traditions collide: northern American, southern American and European (mainly Dutch). It was clear that each tradition had developed its own set of truisms with honest and ardent believers. As a result, implementation and change proved difficult and slower than expected. Not only when it comes to projects knowledge creation seems to matter. As a portfolio GP in several countries – with a Dutch and English registration – I noticed how practices, consultations and contexts vary and outcomes from consultations strongly depend on local interactions between doctors and patients.

Meanwhile I co-founded Generation Next, a think-tank / network of young GPs aiming to look beyond the patient-doctor relationship and bring change in healthcare using an open minded, innovative and friendly approach. In 2008 we noticed that we needed a dedicated network to bring individual practitioners together and share knowledge online wherever they were. This resulted eventually in Dutch National GP Association (LHV) and Dutch College of General Practitioners (NHG, Dutch NICE equivalent) developing HAWeb.nl, a virtual social network with all Dutch GPs as its members. Another start-up became Platform Medical Leadership supported by the Royal Dutch Medical Association. Its goal is to embed leadership in the whole medical profession as well as science. My role as a board member includes building a Dutch competency framework, keeping up relations with international partner organisations such as faculty

of medical leadership in the UK and supporting research on medical leadership in the Netherlands.

While developing these movements, and working in clinical practice, I learned how improvements in healthcare are not only a matter of science but of people too. Influenced in particular by the work of Marc Berg and Thomas Kuhn, I moved to London in 2011 to work with Professor Trisha Greenhalgh at Queen Mary University of London. A National Institute of Health Research In-Practice Research Fellowship helped me to gain a more innovative view on knowledge dissemination, research methods and epistemology while bringing public healthcare and general practice together.

In 2014 I followed an international research training program at Harvard which helped to improve my understanding and skills of randomised controlled trial design and management. Also, I became a standing committee member for a clinical guideline update group at NICE. In 2016 I was awarded a Scientia research fellowship at the University of Oslo, Norway to work with Professors Eivind Engebretsen and Kristin Heggen on clinical guideline development and EBM as a concept. In the meantime, I started work as a coordinator of the academic primary care network in Amsterdam, reviewing research proposals and data enquiries. In all my roles and contexts, it has become very apparent that evidence, knowledge and truisms vary greatly, even when based on the same international evidence or even guidance.

1.5 Structure of this thesis

The remainder of this thesis is structured as follows. In chapter 2 I set the stage to explore the tensions between EBM guidelines and mindlines. First, I develop an ontological and epistemological approach to analyse EBM as a Kuhnian paradigm. Then I focus on views regarding the evolution and shifting of scientific paradigms. Finally, I briefly address current omissions in the epistemology of EBM in order to develop the framework used for analysis in this thesis.

Chapter 3 introduces the methodology and methods behind the parallel parts of this research project:

Secondary research: Theoretical work which was inspired by the principles and quality standards of meta-narrative review [12] and includes a literature review on mindlines, collaborative work on epistemological topics and seminars convened at the Universities of Oxford and Oslo in which researchers and experts identified and discussed key topics, papers and articles.

Primary research which consist of two empirical studies:

- Ethnographic study of the interactions in virtual social network groups of clinicians, looking at how mindlines emerge and are shaped.
- Ethnographic study of guideline development panels to explore the process by which they synthesise knowledge into guidelines and explore how insights from mindlines might be incorporated into their work.

Chapters 4, 5, 6 and 7 present the results from the secondary, theoretical research. This part of the thesis was ongoing throughout the project as publications and chapters were being prepared. It was inspired by the principles of meta-narrative reviewing [12] and appropriate review standards and processes [13]. This part of the research is about bringing together literature from research traditions [14] that cover empirical evidence and theories on knowledge creation and was influenced by discussions from seminars where guideline developers, researchers and experts contributed to, and commented on, the identification of key insights and literature in their field.

In Chapter 4 I reflect on contemporary understandings of knowledge based on a systematic review of the concept of mindlines. This study builds on Gabbay and Le May's naturalistic study of the development and sharing of complex practice-relevant knowledge in general practice [10,15], where they introduced mindlines as a combination of explicit and tacit knowledge which are shared among social groups and which reinforce norms of good practice, but which are also fluid, dynamic and constantly evolving.

I show how mindlines challenge our assumptions about evidence-based knowledge creation, how they require us to re-conceptualise what knowledge is, how they question the criteria we use to designate a component of a guideline as based on 'strong', 'medium' or 'weak' evidence and who should decide what counts as 'valid' knowledge and why [16].

Chapter 5 introduces a framing of EBM as an evolving movement that many called a paradigm shift from the traditional authority of doctors to a modern, authority of rational and scientific knowledge [17–20].

Drawing on 'We have never been modern' [21], an influential work on modernity by French anthropologist and philosopher Bruno Latour, I set out to build an understanding of past and current developments in the discourse of EBM, such as ALLTrials campaign [22,23], 'Shared decision making' [24], 'person centred care'[25] and Option grids [26]. I explain how EBM is largely engaged in a practice of purification to deliver scientific knowledge and what challenges it encounters in the process. In this chapter I address the Implications for clinical reasoning and provide preliminary principles for conceptualizing EBHC as a "situated practice".

Chapter 6 focuses on the fundamental challenges of guideline development. I present collaborative work done with the guidelines international network working group on appraising and including different types of knowledge and reasoning (GIN AID working group). It addresses the challenges that guideline makers face while producing guidelines and develops an epistemological view on these issues.

First it addresses the purposes of guideline production, elaborating on the support for decision making. Then it expands on the so-called problem of induction as introduced by Hume in 1739 [27]. It describes the RCT as an evasion of this fundamental problem and questions the dominance of RCTs in EBHC. Drawing on work by philosopher Ian Hacking this chapter explores alternative ways to evade the problem of induction [28]

for instance making inferences by updating beliefs leading to the best explanation [29], anti-inductivist falsification by Popper [30] and a logic of care by Annemarie Mol [31]

Chapter 7 explores the boundaries of valid and meaningful knowledge. The concept of mindlines and the abundant ways of reasoning may suggest that any kind of knowledge goes. This chapter addresses the concepts of bias, validity, truth and reality to explain what makes knowledge meaningful.

I reflect on how theories of truth operate in EBM/EBHC operate hidden in the central concept of bias. Drawing on work by philosopher Isabelle Stengers [32] I show how bias is in a productive and unproductive conjunction with truth and address the need for more inclusive theories on how to be 'right' in everyday healthcare.

Chapters 8 and 9 present the results from the two primary research studies. Chapter 2 introduced the overall approach, aims and methods: a mixed method qualitative research project, consisting of a digital ethnography of large virtual knowledge networks of physicians and an ethnography of guidelines panels.

Chapter 8 focusses on describing the nature and exchange of knowledge in the studied virtual networks. It addresses the results from the digital ethnography of the 3 online communities in UK, NL and Norway, where the main aim was to demonstrate and characterise the mindlines on such networks. In this study virtual networks are viewed as 'knowledge artifacts', not so much as carriers of objective knowledge, but as "collaboratively created [inscribed artifacts], maintained and used to support

knowledge-oriented social processes (among which knowledge creation and exploitation, collaborative problem solving and decision making) within or across communities of practice" [33,34]. This study focuses on statins on the basis of the large number of posts on this topic due to fierce debates in the medical community on this topic and also to allow comparison across networks.

Chapter 9 addresses the findings and conclusions from the complete ethnography guideline panels in UK, the Netherlands and Norway. It addresses how guideline panels are influenced by different perspectives on epistemic hierarchies and approaches to uncertainty.

The final chapter provides a summary of the findings and arguments advanced during my project. In this chapter I draw all findings from empirical data and theories about clinical knowledge creation together and how they address the research question. I revisit and reflect on the methods used and address the limitations of this thesis. I then proceed to consider the wider the implications of my research for guideline development, and teaching EBM. Finally, I provide recommendations and suggestions for further research.

My intention is to further expand the fundamental understanding of the concept of EBM, and to enrich the theorization of the notion of mindlines in clinical knowledge development, especially how they emerge and get refined through group interaction. Furthermore, I reflect on how to expand the tools we use to produce meaningful knowledge, provide insights into how to overcome the barriers that guideline

development panels face incorporating a broad range of knowledge sources into their recommendations, and how to align guideline processes with the needs of the wider medical community in new and innovative ways, including the use of virtual technology.

Chapter 2 Scope and philosophical approach

2.1 Introduction

In this chapter I set out the philosophical approach and scope to address my research question: how the philosophy of knowledge helps us to explore the tensions between guidelines and mindlines. I start by drawing on work by philosopher of science Thomas Kuhn [14] to develop an ontology of EBM as a scientific paradigm, where anomalies, pioneers and retooling may herald a shift towards a new paradigm. I then give an overview of the literature on the epistemology of EBM, in order to develop a philosophical framework of evidence-based decision making and its components as the central theme of the EBM paradigm. I will use this lens to analyse the literature and empirical data to find anomalies and possibilities of retooling.

2.2 EBM as a paradigm

When considering how to produce meaningful and valid knowledge, it is useful to draw on philosophical notions of the nature of reality (ontology) and how scientists might gain knowledge about that reality (epistemology). Somewhat surprisingly perhaps, much of mainstream EBM was based on a set of philosophical assumptions that went unquestioned – for example, that it was possible, through rigorous empirical study, to reach firm truths about key healthcare problems; that these truths would be (to a large extent) universal and transferable; and that applying methodological principles and rules (for example, the hierarchy of evidence) would assure the quality of this search for

truth. EBM's stated goal was to build this universal knowledge base (and preferably as soon as possible)[20].

Yet there are legitimate questions to be asked about the epistemological foundations of EBM. For example, as Steup puts it "What are the necessary and sufficient conditions of knowledge? What are its sources? What is its structure, and what are its limits? [..] How we are to understand the concept of justification? What makes justified beliefs justified?"[35].

In focusing on the *epistemology* of EBM, I place significant emphasis on the ways in which knowledge is generated. This emphasis, I argue, is of more fundamental importance in studying EBM than particular historical events or the evolution of structures related to the EBM movement. Admittedly, such perspectives would help to describe whole system mechanisms, that might be able to predict new understandings and roles EBM will have in society. However, it gives insufficient means to see what options individual clinicians and patients have in day-to-day practice. As a clinician this is where my main interest lies: how to improve EBM in order for it to work for the single case scenario, an individual, unique patient. For this reason, I choose an epistemic interpretation to explore the tensions between EBM guidelines and mindlines.

For this purpose, I will use the concept of the scientific paradigm as conceptualised by Thomas Kuhn. He defines paradigms as "universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners."[14] He argues that science develops through a succession of paradigms

using many historical examples of theories that were once dominant but have eventually been dismissed. He shows for example how Aristotelian dynamics, phlogiston chemistry, and Newtonian physics helped to explain certain phenomena, were successful in doing so, but over time were discarded and replaced by newer theories.

Thomas Kuhn is considered one of the most influential philosophers of the twentieth century, whose work drew philosophy of science closer to the history of science [36]. His book 'The Structure of Scientific Revolutions' [14], published in 1962, inspired a social science of science itself (i.e. sociology of scientific knowledge, see Sismondo [37]). Kuhn makes clear that he struggled with how other scholars understood his concept of a paradigm in an elaborate postscript added in 1969 to the book [14]. His work has sometimes been interpreted as introducing a constructivist (the notion that knowledge is socially constructed) or relativist (the notion that any explanation is as good as any other) perspective on science, even though Kuhn sought to distance himself from mainstream social constructivism [36] and argued strongly against relativism– for example, he offered criteria for good scientific theories [38]. He opposed critics who accused him of making theory choice "a matter of mob psychology"[38]. Kuhn's constructivism has been described as semantic dependence. In this view, scientific understanding is dependent on the paradigm, as a kind of scientific language or lexicon. It is like attaching new meanings to a word like 'planet' to make more useful distinctions. This is to be differentiated from the dependence on the beliefs of groups of persons in a society as in social constructivism, where nature is viewed as somehow being dependent on us [39].

As a response to his critics, Kuhn describes a narrow and broad definition of paradigms in the postscript:

"what the members of a scientific community share, and, conversely, a scientific community consists of men [sic] who share a paradigm"[14] *and can be used in two senses; "On the one hand, it stands for the entire constellation of beliefs, values, techniques, and so on shared by the members of a given community. On the other, it denotes one sort of element in that constellation, the concrete puzzle-solutions which, employed as models or examples, can replace explicit rules as a basis for the solution of the remaining puzzles of normal science."*[14]

Kuhn writes that he was inspired by Ludwik Fleck's *Entstehung und Entwicklung einer wissenschaftlichen Tatsache*, published in 1935 [14]. He later wrote the preface of its translation: Genesis and Development of a Scientific Fact [40]. Fleck, a Polish and later Israeli microbiologist and epistemologist who worked on vaccines, argues how scientific truths are always framed by the thought style (Denkstil) of the collective of observers, i.e. researchers (Denkkollektiv) [41–43]. Using a detailed historical analysis on how views on syphilis changed over time, he explains how for example 16th century scientists were not in a position to replace "the mystical-ethical view of syphilis for one based on natural science and pathogenesis."[40] Fleck sees a "cultural historical dependence", a "stylistic bond" of interconnecting concepts, the thought style which is upheld by a community. He writes:

"If we define 'thought collective' as a community of persons mutually exchanging ideas or maintaining intellectual interaction, we will find by implication that it also provides the special "carrier "for the historical development of any field of thought, as well as for the given stock of knowledge and level of culture. This we have designated thought style"[40]

He further explicates the idea of collectives as both individual and collectively held:

"Although the thought collective consists of individuals, it is not simply the aggregate sum of them. The individual within the collective is never, or hardly ever, conscious of the prevailing thought style, which almost always exerts an absolutely compulsive force upon his thinking and with which it is not possible to be at variance."[40]

Flecks ideas permeate not only Kuhn's work, but can also be seen in for instance Foucault's 'episteme' ("the 'apparatus' which makes possible the separation, not of tile true from the false, but of what may from what may not be characterised as scientific."[44]) and Gabbay and Le May's 'mindlines' ("collectively reinforced, internalised, tacit guidelines"[10]). The latter concept will play a significant role in this thesis and will be discussed in more detail in Chapter 4.

What sets Kuhn's work apart from Fleck's is his focus on *how* paradigms shift. According to Kuhn, theories compete for existence throughout history. Most scientists are usually engaged in 'normal science', which means that they try to "solve the puzzles"[14] within the boundaries of an existing scientific paradigm. New, odd observations are first fitted

in the old paradigm until under certain external conditions these inconsistencies give rise to a "crisis"[14] and a new alternative theory is adopted: a scientific revolution.

I will discuss how I intend to use Kuhn's ideas on the succession of paradigms in this thesis. But first I address the question whether it is reasonable to see EBM as a paradigm in Kuhnian terms.

A significant argument in favour of this position is that EBM pioneers explicitly drew on work by Thomas Kuhn to introduce EBM as a concept. In an article in JAMA in 1992 the EBM Working Group writes:

> *"A new paradigm is emerging. EBM de-emphasises intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research."* [20]

The authors explain this claim by paraphrasing Kuhn's work as follows:

> "*Thomas Kuhn described scientific paradigms as ways of looking at the world that define both the problems that can legitimately be addressed and the range of admissible evidence that may bear on their solution. When defects in an existing paradigm accumulate to the extent that the paradigm is no longer tenable, the paradigm is challenged and replaced by a new way of looking at the world. Medical practice is changing and the change, which involves using the medical*

literature more effectively in guiding medical practice, is profound enough that it can appropriately be called a paradigm shift." [20]

Some critics have argued that the EBM movement does not have the key features of Kuhnian paradigms [45,46]. They for instance accuse the early EBM movement for example, of failing to adequately define the previous paradigm. However, at no point does Kuhn state that scientists introducing new paradigms have to do this. A valid criticism could be that the EBM pioneers viewed paradigm shift as a methodological approach ("first, define the old paradigm; then, take the following steps to replace the paradigm"). But Kuhn's 'social science of science' is not a set of normative methodological steps but a naturalistic view of how science unfolds whether or not scientists are aware of it.

Yet even though some authors continue to contest whether EBM satisfies any of Kuhn's definitions of a paradigm [47], for a new generation of doctors EBM is the current dominant paradigm in the medical world. EBM behaves as a paradigm in Kuhnian terms in the sense of being the current dominant set of "beliefs, values, techniques, and so on shared by the members of a given community"[14]. In this paradigm RCTs function as the ultimate puzzle-solution to problems encountered in healthcare. Tools for the structured analysis of published studies, like GRADE [48], shape guideline development. And the medical community largely promotes EBM and medical students are trained in the ways of EBM [49].

2.3 Shifting paradigms

If EBM can be considered a paradigm (the result of a shift from an older paradigm), then to demonstrate and explore tensions in evidence-based knowledge development would mean to find the indicators that signal the shape of a new paradigm that will sooner or later replace EBM. Kuhn's work offers hints of the kinds of signals to look for.

A first strategy would be to look for anomalies with the potential to cause a revolution under certain external conditions. Every so often such a new finding emerges, or a criticism is formed in such a way that it disrupts the consistency of an established paradigm. According to Kuhn, most such findings are initially fitted into the old paradigm or disposed of as errors or anomalies, but some are not easily explained away, and may then accrue a following with little input from resources, like a spark. New practices, beliefs and techniques that could form the basis of a new paradigm are able to develop further taking up the old resources. They initially emerge in a wide variety of forms with little consistency between them (the so-called "pre-paradigmatic phase"). Then comes a phase where several significant theories compete. New resources are invested; new tools are developed. The consistency of the new ideas is tested; theories and techniques are researched and explored in empirical studies. A dominant paradigm slowly emerges again. Gradually, a set of accepted principles, rules and approaches becomes standard – what Kuhn called the "normal science" of an established paradigm.

Kuhn argued that an established scientific paradigm enables scientists to achieve a high level of sophistication in addressing what have been agreed as the key problems and challenges in that paradigm:

> *"During the period when the paradigm is successful, the profession will have solved problems that its members could scarcely have imagined and would never have undertaken without commitment to the paradigm. And at least part of that achievement always proves to be permanent."* [14]

But Kuhn also observed that when the paradigm is at its peak, the tremendous precision, concentration of resources and depth of scope accrued will not only help to solve problems, but also identify new boundaries and difficulties. Advanced tools and techniques unearth new findings that do not completely fit the new paradigm.

A second indication of a looming paradigm shift are pioneers outside of the established paradigm. Kuhn writes how scientists who have made a career out of the now old paradigm will seem to stubbornly resist a new paradigm. This resistance, he argued, is not unscientific – it is how science needs to work to build robust paradigms capable of going the distance and find new boundaries. In general, therefore, individual scientists who have led one paradigm are rarely the ones who lead a change in paradigm. Instead a new, alternative (sometimes but not necessarily younger) generation of scientists emerges who build the new paradigm. Kuhn cites Charles Darwin and Max Planck:

"Darwin, in a particularly perceptive passage at the end of his Origin of Species, wrote: 'Although I am fully convinced of the truth of the views given in this volume ..., I by no means expect to convince experienced naturalists whose minds are stocked with a multitude of facts all viewed, during a long course of years, from a point of view directly opposite to mine. . . . [B]ut I look with confidence to the future,—to young and rising naturalists, who will be able to view both sides of the question with impartiality.' And Max Planck, surveying his own career in his Scientific Autobiography, sadly remarked that 'a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.'[14]*"*

A third indication of paradigm shift is what Kuhn calls 'retooling'. When one paradigm or theory dominates, all resources, skills, money and people are devoted to the techniques and processes that are required to the expense of other tools. Going beyond a paradigm like EBM is costly because it not only requires new theories and methods, it also needs re-tooling. Kuhn writes:

"Philosophers of science have repeatedly demonstrated that more than one theoretical construction can always be placed upon a given collection of data. History of science indicates that, particularly in the early developmental stages of a new paradigm, it is not even very difficult to invent such alternates. But that invention of alternates is just what scientists seldom undertake except during the pre-paradigm stage of their science's development and at very special occasions

during its subsequent evolution. So long as the tools a paradigm supplies continue to prove capable of solving the problems it defines, science moves fastest and penetrates most deeply through confident employment of those tools. The reason is clear. As in manufacture so in science—retooling is an extravagance to be reserved for the occasion that demands it. The significance of crises is the indication they provide that an occasion for retooling has arrived."[14]

For the scope of this research I choose to focus mainly on uncovering anomalies and needs for retooling, and not so much on finding pioneers of new paradigms. While doing so I will take into account that criticism of Kuhn's views on how paradigms change has subsequently led to a more nuanced view on paradigms shifts. First, transitions from paradigm to paradigm may be more subdued and less revolutionary. They can also change slowly over time through normal science [50]. Moreover, it does not always require anomalies to cause a scientific revolution. New insights from findings from normal science can open up entire new fields of research without changing the dominant paradigm. For example, the discovery of the double helix structure of DNA solved the issue how genetic information could duplicate [36]. Likewise, on closer reading the above mentioned evidence-based working group kept the importance of clinical experience from the old paradigm in the new EBM paradigm [20]. They wrote as the first point of the new paradigm: "Clinical experience and the development of clinical instincts (particularly with respect to diagnosis) are crucial and necessary part of becoming a competent physician. Many aspect of clinical practice cannot, or will not, ever be adequately tested." [20] This point only differs from the first point they describe as the old paradigm in that it is deemed *insufficient*: systematic observations need to be

added. Searching for tensions between guidelines and mindlines as indicators for how EBM reaches its limitations may therefore not result in finding any extraordinary new practices or theories as they may well emerge quietly and through nuanced adjustments through the course of practicing it.

2.4 The epistemology of EBM

In this paragraph I turn to what kind of anomalies and tools are of interest to help us explore the tensions in evidence-based knowledge development. The question is what could currently be considered the a priori, thought style or apparatus that precedes any knowing in EBM?

As mentioned earlier, EBM pioneers did not start out with an elaborate debate of underlying philosophical assumptions. However, a body of literature on EBM epistemology has emerged over the past decades. A significant portion of this literature was developed as a critique of EBM, that according to a systematic review could broadly be categorised into "a reliance on empiricism, a narrow definition of evidence, lack of evidence of efficacy, limited usefulness for individual patients, and threats to the autonomy of the doctor/patient relationship" [5]. This summary of positions provides little insight in my view and I will try to sketch out what epistemology is suggested in the literature, drawing a more detailed analysis of work by Mark Tonelli, Benjamin Djulbegovic, Jeremy Howick and Miriam Solomon instead. I acknowledge that this approach may put too much emphasis on these authors. I selected these authors as they

appeared to me as having a more significant influence on the epistemological debate around EBM throughout the years I worked on this topic. Although all are professionally engaged with EBM, I view Djulbegovic and Howick as examples of authors who are more inclined to defend EBM, whereas Tonelli and Solomon exemplify those who take a more critical approach. I first provide an outline of their work and then provide a brief analysis.

Mark Tonelli is one of the early critics who specifically questions the epistemology of EBM. He writes in his paper The Philosophical Limits of Evidence from 1998:

"EBM deals with the issue of knowledge in medicine, defining optimal ways to develop knowledge and describing hierarchies of medical evidence. Therefore, it represents a school of medical epistemology." [4]

In his analysis Tonelli notices that the early EBM movement sought to limit the use of knowledge that was not derived from clinical trials. Later, the focus of EBM moved towards integrating different kinds of knowledge, including patient preferences, pathophysiological rationale and clinical expertise. Thereby, Tonelli argues, it reinstated some of their value once again, albeit not on the same standing. What makes EBM a new paradigm is that it differentiates itself from alternatives such as value-based, physiologically-based or experientially-based by making evidence from RCTs take "general priority" in most (but not all) individual cases [4]. Tonelli notes that the epistemological challenges that are not addressed by EBM are firstly, how to integrate different kinds of evidence and secondly, how to understand in what circumstances other knowledge should have priority [4].

Tonelli goes on to argue that EBM is often envisioned to close the gap between clinical research and clinical practice but is yet unable to do so for philosophical reasons. He contends that it may work for public health but not for the individual patient. The reason is that the gap consists of individual variety from the average and cannot be fully quantified. Tonelli argues that randomization to reduce the effect of non-quantifiable factors and elicit the impact of quantifiable factors on individuals won't do. Non-quantifiable factors, which could for instance be known from clinical expertise, are relevant to clinical reality. In Tonelli's view the problem of quantifiability of clinical reality is a major epistemological challenge for EBM. This extends to how EBM deals with patient values. Values answer ethical questions which cannot be derived from facts. EBM would recognise the importance of values but attempts to quantify them in "awkward' concepts as for in for instance patient utilities [4].

Tonelli contends that the risks of not acknowledging the limitations of EBM include limiting the individual to his or her quantifiable aspects (disregarding the importance of "physical, emotional and spiritual aspects of illness in the healing process"), secondly a shift away from the individual to groups of patients (contrasting the traditional focus on the individual in medicine) and thirdly, the denial of the complexity of clinical judgment. Regarding the clinical judgement, Tonelli is critical of models of clinical decision making, which are not adequate to describe actual clinical reasoning by experts, most importantly because they lack the ability to capture tacit knowledge and patient values.

"The recognition of the value of other forms of medical knowledge and reasoning by EBM represents an important step in developing a complete and coherent account of the unique undertaking that is clinical medicine." [4]

Benjamin Djulbegovic wrote a paper with Gordon Guyatt who coined the term EBM [51], and Richard Ashcroft in which they challenge various criticisms of EBM [52]. They discuss how EBM relates to philosophical theories on evidence, knowledge and science and feel that much of the critique is misplaced. In their paper they address several partly related issues.

First, they consider the difficulty of defining evidence. They state that the least controversial view is to see evidence as grounds or justification for true belief. In this view knowledge is true belief, but evidence is not necessarily true. Central to EBM would be its emphasis on uncertainty and scepticism. Evidence is then used in several modes: to support rational decision for populations and individuals; as a guide to truth in hypothesis testing and as an arbiter to settle disputes and reduce variation in clinical practice.

Secondly, they discuss the reliability of measurements and observations. They consider how EBM deals with the question whether we can go beyond observable events to make valid yet unobservable predictions. They regard EBM as adhering to a logical-positivist tradition through its belief that only reliable empirical observation using a set of rules separates truth from untruths. As such, EBM would not make claims about how the world is, but merely to predict future empirical observations. It is not supposing any

unobservable mechanisms underlying these observations, it is only interested in the accuracy of them. The authors warn however that EBM should not be equated with logical-positivism, as the latter would find that anything that would be observable in principle but not observable all the time nonsensical, while EBM would be more lenient and pragmatic about this. They then write how EBM puts emphasis on rigorous standards and checklists in an effort to develop a universal set of rules to reach accurate inferences based on empirical observations. But they go on to play down this focus on rigor by arguing that ultimately EBM is not about the validity of scientific claims but rather about providing rules that can be used in clinical practice, where many aspects cannot be quantified. On a side note they grant that qualitative research challenges the view that what cannot be quantified is unobservable. However, they find that qualitative research is merely a different kind of empirical evidence which does not provide better access to reality [52].

In the third part of the paper, they examine EBM from three epistemological points they call inductivism, falsificationism and explanationism. They see inductivism in EBM as reasoning from a sample (like symptoms or clinical trials) to more generalised statements (like a diagnosis or a guideline recommendation). Falsificationism is endorsed by EBM as it is engaged in hypothesis testing as a method of verification of scientific statements. However, they claim that EBM denounces a binary approach such as a strict application of the p-value as a threshold for truth. Instead it would acknowledge that observations are theory laden in the sense that EBM is not consistent in proclaiming the neutrality of evidence, for instance in the case of homeopathy, where results from good scientific designs are more doubted than in less controversial

disciplines. The authors see signs of explanationism (or holism) in EBM as it embraces a view that new scientific evidence should cohere with existing explanations, knowledge and beliefs. They argue that EBM is not about new philosophical theory and does not "have a rigorous epistemological stance." Instead, EBM draws on many traditions of philosophical theories of science and should be viewed as "a structure for optimal clinic practice" [52].

In the final part of the paper Djulbegovic et al. address the nature of scientific methods. Considering EBM as a Kuhnian paradigm, they argue that its rise reflected the development of public requirements regarding the scientific method such as medicine testing by the FDA. Some have regarded this as a democratic force, allowing anyone to challenge claims by authoritive figures such as professors, provided they understood the scientific method standards. On the other hand, EBM is seen as an authoritarian force: the strict standards, checklists and hierarchies of evidence have stifled scientific debate. The authors argue that within EBM there is acknowledgement that a universal set of rules for evidence may not be feasible. Yet, they uphold that EBM embraces a strong belief that a 'wise man' bases his convictions on evidence [52].

They acknowledge that EBM has promoted reductionism in the scientific method, a "neglect of postmodernist ideas of social importance of science activity" and of not providing a theoretical framework for successful decision making [52]. However, they claim that EBM has responded to criticisms of the limitations of basing inferences on group observations by placing more emphasis on the role of patients in decision making.

They conclude that EBM is not a new philosophical theory, but rather a "continuously evolving heuristic structure for optimizing clinical practice" [52]. Nonetheless they write in their conclusion:

> *"The key challenge that needs to be tackled is how to elucidate a process of optimal decision-making. This includes the difficult task of defining rational decision- making for individuals and society and the application of evidence obtained from group observations at the patient bedside"* [52].

Jeremy Howick's book 'The Philosophy of EBM' from 2011 is an attempt to describe and defend an epistemology underlying EBM. Howick contends that "in spite of evolving characterizations, the EBM view that comparative clinical studies, preferably (systematic reviews of) randomized trials, provide more telling evidence for therapeutic effects than mechanistic reasoning and clinical expertise has remained constant." [19] What counts as good evidence is fundamentally related to the purpose of EBM, that is to support decision making in everyday practice [53].

Howick finds this EBM philosophy reasonable and acceptable. However, he proposes some changes to what counts as best evidence in EBM. He discusses several specific issues of clinical studies regarding masking, randomization and the use of placebos. For instance, in an attempt to solve the problem that the most effective treatments in medicine (like the Heimlich manoeuvre and insulin for diabetes) are not supported by randomized controlled trials, he argues that randomization is not always required. He

poses that observational studies should be considered especially when the effect size is so large that it outweighs confounders. He also argues that EBM should allow mechanistic reasoning in some cases as a valuable form of evidence within the hierarchy of evidence to make an inference, mainly when the mechanistic reasoning is not incomplete and of high quality [19].

His most radical suggestion however is to denounce expert judgement as a source to make an inference. Instead of EBM being about integrating best evidence with clinical expertise and patient values, putting them on equal footing, he would say that EBM requires clinical expertise for producing, interpreting and integrating evidence with patient values and context. Clinical expertise is in his view only supportive, but never reliable to make an inference [19].

Howick's central position is that observational studies, mechanistic reasoning and clinical expertise are of a lesser standing and EBM is indeed right to prefer evidence from RCTs above other types of evidence in most cases. He supports this with an elaborate exposition of systematic reviews from Meehl [54] and Grove [55] that show that statistical prediction rules outperform expert judgement in most situations. Meehl and Grove did most of their work on non-clinical prediction rules, but a clinical example would for instance be the Ottawa ankle rule [56], which is used to predict whether an x-ray is required in a patient suspected of a fractured ankle. In Howick's view, the abundance and diversity of these studies all pointing in the same direction (albeit their small sample sizes) should convince anyone to see statistical prediction as superior to

clinical judgement. In contrast, tacit knowledge and skills of an expert would only be needed to integrate evidence, but they should not be used for prediction.

The future of EBM is about dealing with new biases and extending its scope, Howick finds. The first is conflict of interest: the problem that industry sponsored trials more often show effectiveness of an intervention than those that are not sponsored. This could be solved with a 'new critical appraisal level' that would include blinding data analysts and trial registration, Howick suggests [19]. Furthermore, as there are more ways to improve health beyond clinical treatment, EBM movement should broaden its methodologies to support this.

In her 2015 book, 'Making Medical Knowledge' [57] but also in earlier work [58], **Miriam Solomon** develops a more comprehensive epistemology. She criticises EBM for only being an incomplete epistemology. Instead of a need to absorb different kinds of knowledge in the classic hierarchy of evidence as suggested by Howick and others [19,59,60], Solomon argues that EBM is just one of many ways to make knowledge that stand in their own right [57]. She discusses three alternative methods in detail: consensus conferences, translational medicine and narrative methods. Like EBM, each with their own advantages and caveats.

In her view EBM stands on an empiric tradition, bringing in rigour, verification and control for biases. However, it is more fallible than anticipated for various reasons in practice, such as publication bias, conflicts of interest and other factors not yet fully understood. Consensus conferences stress rhetoric and trust of authorities as means to

produce knowledge. She points out that it is telling that the GRADE methods to appraise evidence for systematic reviews and guideline development, were created in consensus meetings. The downside of consensus is that it may not be able to introduce innovations very well when they are surrounded by controversy. Translational medicine (more specific, T1: bringing findings from the laboratory research to a clinical trial), is based on informal experimentation, building on knowledge of mechanisms and hypothesis testing. A significant disadvantage is that this method often leads to knowledge that turns out to be unsuccessful in practice. Narrative medicine is interested in the knowledge that develops in the relationship between clinician and patients as well as what can be learned from sharing case-studies. It brings in the act of witnessing, the power of empathy and creating meaning as elements of medical knowledge. But it can also bring the consultation wrong information in the form of incorrect recollections, confabulations and under-and overemphasis of events to build a coherent story [57].

Solomon explicitly takes a perspective that is pluralist in nature. There are many methods, that are not mutually exclusive, that often collide, and they have their own merits and downsides. She remains appreciative of all of them.

"EBM gives a set of formal techniques for evaluating the effectiveness of clinical interventions. The techniques are powerful, especially when evaluating interventions that offer incremental advantages to current standards of care, and especially when the determination of success has subjective elements." [58]

Looking at the future of EBM, she notices that new contenders will be around eventually. She sees personal medicine as one of those new methods that comes into fashion [57]. It is, in her words, a confusing method that brings together two quite distinct perspectives: precision treatments based on the genetic makeup of subgroups (sometimes just one person) and a view that every patient requires a unique, personal and intimate approach [57].

Analysis

Overall, Tonelli, Djulbegovic et al., Howick, Solomon and others indicate that the early EBM movement paradigm should be defined by its claim that evidence from RCTs should trump dominant pathophysiological theories and other kinds of evidence in most cases [19,46,61]. This is exemplified by the development and adherence to the hierarchy of evidence, with systematic reviews of clinical trials in the top, and expert opinion on the bottom.

For Howick and Djulbegovic et al., making clinical decisions for individual patient is another central part of the EBM paradigm. Solomon differs slightly, in the sense that apart from the narrative method, she does not go into much detail how these other methods of creating knowledge relate to medical decision making for individual patients. Tonelli's is more concerned that EBM shifts a focus on the individual towards society. This is noteworthy as the definition of EBM given by Sackett et al. does define a strong focus on decision making for individual patients. In 1996 they were responding to critics explaining what EBM is and what it is not:

> *"Evidence-based medicine is the conscientious, explicit, and judicious* use *of current best evidence in making decisions about the care of individual patients."* [3]

The previously mentioned definition of the EBM working group [20], which dated a few years earlier, which says EBM is about de-emphasising intuition, unsystematic clinical experience, and pathophysiologic rationale, also urges to take evidence from clinical research into account when making decisions. Yet, neither definitions nor the papers they come from articulate what that decision-making entails. As Djulbegovic et al. point out, a tension remains in the epistemology of EBM how and whether it actually achieves to make good predictions for individual patients.

Tonelli, Solomon and Howick are critical of the hierarchy of evidence, yet at the same time continue to consider the techniques used in EBM as 'powerful' [57] means to reduce bias. Solomon and Tonelli are not clear on what epistemological grounds this is the case. Howick's suggested epistemology of EBM tries to defend the hierarchy, but his arguments are shaky. He contends that findings from observational studies and mechanistic reasoning can be on equal footing with evidence from randomized trials in exceptional cases (when other evidence for effectiveness is so convincing that it trumps a clinical trial). Yet it is not clear what is required to make a case 'exceptional' [53]. A major blow to Howick's central position however comes from the EBM movement itself. In a 2014 systematic review of clinical prediction rules (CPRs) the authors noted something that the studies on (mostly non-clinical) prediction rules versus expert judgement by Meehl and Grove had overlooked. Better prediction is not just about

being more often right, it is also about being less often wrong. This makes the dominance of statistical prediction rules less convincing. The authors write:

"In the limited studies to date, CPRs are rarely superior to clinical judgment and there is generally a trade-off between the proportion classified as not having disease and the proportion of missed diagnoses. Differences between the two methods of judgment are likely the result of different diagnostic thresholds for positivity. Which is the preferred judgment method for a particular clinical condition depends on the relative benefits and harms of true positive and false positive diagnoses." [62]

Many scholars have concluded that other kinds of knowledge cannot be dismissed [2,59,63]. Djulbegovic et al. argue that this is a result of a logical positivist stance within EBM. What is lacking is an exploration and discussion of whether the underlying assumptions in EBM that drives a preference for statistical prediction are sustainable if the aim is to improve individual decision making in clinical practice.

Solomon's epistemology provides the most comprehensive exposition of knowledge creation processes in the many disciplines of medicine. However, they are presented as nearly parallel practices. As a result, she sees mechanistic reasoning outside of the hierarchy of evidence, where Howick explicitly keeps it in. Howick further dismisses the need for tacit knowledge as he considers it as of no importance to make good predictions, while Tonelli, Solomon and others [64] stress that it is a significant and necessary kind of knowing. This underlines an unsolved issue in the epistemology of EBM regarding the place and role of different kinds of knowing.

Furthermore, Tonelli, Solomon and Howick present no analysis of EBM's assumptions regarding bias, besides statements that it should be avoided. There appears to be a general omission in the epistemology of EBM on this subject. This would require a study of notions of truth and reality in healthcare. Tonelli's and Djulbegovic's remark that EBM is unable to fully quantify reality (therefore needing clinical judgment), further calls for an exploration how what we know relates and can be derived from that reality.

Scholars have pointed out that EBM paradigm has an axiological (assumptions about what is of value) dimension too [65,66]. Values refer to how we, as observers, think the world ought to be [65]. Kelly et al write how values influence how we understand reality:

"[..] not as the reflective 'side' of our psychology (in contrast to the rational 'side') but as the essential texturing of everything we perceive, believe and aim for." [65]

Financial value [67] and patient values are commonly addressed in EBM, but rather as an extrinsic form of evidence ready to be integrated. Kelly et al. show how EBM processes (like the choice of topics, what evidence to include and who are considered stakeholders) are highly impacted by the value of utility and efficiency, the belief that the greatest health gain for the greatest number is morally the best approach [65]. A view of facts as inherently value-laden seems to be missing critics note. There appears to be an implicit and often unquestioned assumption that we ought to chase the most effective way to achieve health and that EBM is the way to maximize the likelihood of arriving at the truth [68,69].

In the final part of this chapter I aim to bring all these criticisms and gaps of the EBM paradigm together in a framework for analysis used in this thesis.

2.5 Developing a framework for analysis

First, I take the EBM paradigm's constellation of beliefs, values, techniques and so on to revolve around the definition given by Sackett et al., with an emphasis on the use of evidence from clinical trials to support decision making for individual patients.

Although EBM appears to lack a stable epistemology of decision making, it is a topic extensively researched throughout many disciplines, including mathematics, economics, education, psychology and sociology. Many scholars have addressed particular issues in clinical decision making, including Gerd Gingerenzer [70] on heuristics, Annemarie Mol [31] on the logic of care versus the logic of choice, Katherine Montgomery on the importance of clinical judgement [71], Bernard Lonergan on interpretational operations to integrate knowledge [72] and Nobel prize winners Kahneman and Tversky on cognitive biases [73], to name a few.

In order to get some kind of an epistemological understanding of all these models and research traditions regarding clinical decision making, I choose to largely follow contemporary philosopher Ian Hacking [28] in his book 'Probability and Inductive Logic' by considering decision-making epistemically as ways to make an inference. In this view the clinician and the patient are faced by the philosophical problem of induction [28]. In

short, this problem means that it is impossible to predict the future. It is important to remember as noted above that Sackett's definition of EBM stresses the decision-making for an *individual* patient [3]. I will take this into account by particularly focusing on inferencing in the single case scenario.

In this epistemological view on decision making, the starting position becomes a reality where a clinician is faced with a real problem of an individual real patient (and real organization and systems) in a real context. The end position is the outcome for this patient in a near future reality. For instance, imagine the actual patient a GP will see tomorrow morning at her surgery somewhere in the UK, presenting with a kind of burning sensation in the lower back, that she has never experienced before. Yes, there have been many patients with backaches, but the event of this patient presenting with this sensation at this time, at this place, with this context is unique. Now imagine her actual situation a week from now. Her pain may have subsided, she may have more pain, she may not attend. Her future reality will be very real but tomorrow during the consultation it remains uncertain. She and her GP will be in a process of decision-making tomorrow morning somehow effecting how her future reality a week later will look like. Taking into account the gaps in the epistemology of EBM as discussed above, I take the dimensions of this setup to be:

1) the nature of the knowledge that goes into the decision-making,
2) the way that knowledge is related to reality,
3) the reasoning to make the decision and
4) the nature of current and [possibilities for] future reality.

These four dimensions are not mutually exclusive, and events and statements may be explained in several of them. Yet, assumptions about these dimensions I take to constitute the metaphysical beliefs and values of the EBM paradigm that underlie its practices, tools and techniques. I will address these dimensions of decision-making and their anomalies, new insights, difficulties and criticisms in the theoretical part of this thesis. In the empirical part of this study I aim to look at the nature of these elements and anomalies in reality.

Regarding demonstrating where EBM knowledge development reaches its limits and tension with mindlines emerge when it comes to its *tools* as an indicator of new paradigm shift, I choose to limit the scope of this thesis to guidelines. The number of EBM tools developed over the years is extensive and includes tools to appraise knowledge (like GRADE [48,74]), frameworks to structure review strategies, rules for reporting conflicts of interest [75], clinical pathways and many more. I choose to focus on guidelines not only because their development requires many other tools, but also because they are considered to have a central role in practicing EBM in such extent that they are often equated with it. For instance, in *The Gold Standard,* Timmermans and Berg analyse and discuss the practicalities of the implementation of EBM, protocols and standardisation. They write:

> *"in common medical parlance [EBM] mainly denotes the use of clinical practice guidelines to disseminate proven diagnostic and therapeutic knowledge."* [18]

I will analyse how guidelines are, and could be adjusted, in Kuhnian terms *re-tooled*, fitting paradigms of other beliefs, values and techniques able to cope with the difficulties, criticisms, anomalies, crisis and new insights that EBM faces currently. In the methods chapter I will address how I intend to do this in more detail.

2.6 Conclusion

In this chapter I have laid out the ontological and epistemological approach and scope of this thesis. It sees EBM as a Kuhnian paradigm with particular beliefs, values and techniques revolving around the decision-making support. In order to explore the tensions between guidelines and mindlines I will demonstrate how EBM reaches its limits and explore paradigms that may lie beyond. In work I present later in the thesis I will try to fundamentally challenge current framings of EBM by Howick, Djulbegovic, Solomon and Tonelli. The objective will be to look for anomalies, i.e. alternative beliefs, values and techniques by questioning how guidelines, as the central tool in practicing EBM, could be re-tooled to fit and cope with current criticisms of EBM.

Chapter 3: Methodology and methods

3.1 Methodology

In Chapter 2 I set out the epistemological approach and scope of this thesis. Using Kuhn's work on paradigms to explore the tensions between guidelines and mindlines I developed a philosophical framework with four dimensions. In order to empirically examine these dimensions, I have pursued a methodology that would allow me to look in detail at knowledge development processes in theory as well as in practice. The designs and methods of the projects in this thesis were informed and revised over the course of several years in discussions with many, but mainly through collaboration with my supervisors. In this section I address the general methodology that was chosen for this thesis. In the following sections I detail the specific methods and study designs that were employed.

For the primary and secondary research projects in this thesis I chose a naturalistic empirical approach and interpretive analysis. I intended to undertake a scholarly empirical study that was relevant to practitioners. The aim was not to make general statistical statements about guideline developers or the institutions where they work as a group, nor to describe the frequency or variations of how they work. Rather, the objective was to gain detailed, in-depth data to explore tensions as well as to shape a meaningful comprehension of evidence-based knowledge creation. I consider qualitative methods as the most appropriate means to achieve this as they support a broadly 'interpretivist' position (meaning that they are mostly interested in how the

complexities of the social world are perceived and experienced by humans), provide flexible methods for collection that are sensitive to the phenomena that they intend to investigate (instead of applying stringent abstract categorisations to them) and offer more holistic methods of analysis (that maintain sufficient consideration for detail, complexity and local context) [76]. Furthermore, as sociologist David Bloor writes, the focus of qualitative research "on everyday activities makes it particularly relevant in helping practitioners to think about their working practices" [77].

These common features of qualitative research are essential for my research as they allowed me to (re-)interpret the ideas, social activities and interactions of guideline makers, the processes of guidelines developing institutions, and the documents that are produced throughout, with much detail and in their actual, non-abstract contexts, needed to build a critical understanding of EBM that is consistent with and relevant to everyday practice. I took an interpretivist approach by design, meaning that I assessed my research in terms of reflexivity and rich theorising, rather than a positivist approach, underpinned by a belief that facts are out there, ready to be mined and benchmarked by inter-rater reliability. Most qualitative research is done in the interpretivist tradition [78], and as this thesis is about the fundamentals of knowledge creation, I considered this approach best to develop a meaningful answer to my research question.

An additional reason why I have chosen qualitative methodology for this thesis is the so-called problem of self-authentication of styles of scientific reasoning [79]. As philosopher Ian Hacking points out, different disciplines and branches in the sciences use different ways of thinking and strategies of finding things out. These styles

themselves are self-authenticating in a the sense that they are autonomous without clear foundations: “they do not answer to some other, higher, or deeper, standard of truth and reason than their own” [79]. Quantitative methodology as typically used in EBM, could be considered such a scientific style of reasoning, unable to be criticised by its own way of finding out. In order to evade the problem of self-authentication and enable researching the foundations of reasoning in EBM, I consider using qualitative methods, applied in the interpretivist tradition, to be more appropriate than quantitative methodology.

I began this thesis with an interpretive synthesis of the literature on the concept of ‘mindlines’ to discuss and extend the notion of knowledge in EBM. This secondary research drew on the principles of meta-narrative review [12,80] which I will explain in more detail in chapter 3.2. We selected this approach because it is designed to capture different underlying assumptions about, and approaches to, a concept. Inspired by Kuhn’s work on scientific paradigms (as mentioned in Chapter 2), It allows the synthesis of large bodies of evidence from multiple research traditions, that may have different views on what counts as a legitimate problem, as well as what theories, methods techniques to use, to investigate a similar topic [12]. As such this method provides a holistic program to explore and extend the concept of knowledge in EBM.

For the primary research part of this thesis, I used ethnography (including digital ethnography) to explore tensions between formal and informal knowledge development. The first study was set up as a digital ethnography of the interactions in virtual social network groups of physicians, looking at how informal knowledge is

formed and shaped. In section 3.4 I explain in detail why we chose for the online environment, here I want to clarify why we chose digital ethnography as a method.

Digital ethnography takes many forms including virtual ethnography, internet ethnography, sensory ethnography and hypermedia ethnography, each with a slight difference in application and epistemic assumptions, but akin in their use of digital technologies [81]. These technologies, such as online questionnaires, digital video, social networking websites and blogs, offer social scientists a multitude of new tools to do research with [82]. Digital ethnography can examine the same concepts that researchers in the humanities find useful in any kind of ethnographic research: experiences (affective, sensory, embodied), practices (what people 'do'), personal relationships, social worlds (as a theory of communities), things (and how they are made meaningful), localities, and events; albeit somewhat differently [83]. For instance, whereas conventional ethnography can explore the influence of physical environments on human experiences and action, digital ethnography can do the same for the digital environment.

Pink et al. state five core principles of digital ethnography: there is more than one way to engage with the digital e.g. broadband, smartphones, games, etc. (multiplicity); digital media are part of other, non-digital relationships and activities (non-digital-centric-ness); it is a flexible research design that can be fitted to the specific research question and context (openness); it continuously asks itself how it produces knowledge in a digital world (reflexivity); and it engages in alternative forms of communicating 'beyond the standard written production of academic scholarship' such as ongoing

collaboration and dialogue with the research participants (unorthodoxy) [83]. These characteristics are particularly helpful to efficiently analyse informal medical knowledge interaction and development in very large groups of clinicians with much in-depth detail. For these reasons I found digital ethnography the method of choice for the first primary research project presented in this thesis.

For the second primary research project, which looked at formal knowledge production, I chose conventional ethnography. Guideline production is not purely producing and handling explicit data and documentation, but revolves around meetings with for instance stakeholders, experts and lay people. In order to gain an in-depth understanding of these meetings, conventional ethnography as a method helps to prevent overlooking the 'interpretive skills' of (groups of) guideline developers and others as the study subjects [77]. Furthermore, it not only allowed us to ask explicit questions for instance in the form of interviews as means to investigate, but it also enabled us to observe the actual activities and routines of guideline developers. As such this method provides a way in not only studying the contents of documents produced in the process of guideline development, but also how they are shaped and handled by social actors. Moreover, as a methodological strength, it recognises the researcher's connectedness with the topic of interest. It specifically addresses the tension between scientific detachment and involvement that is often overlooked in other methodology. It offers ways to control and use this tension in the form of participatory as well as observational techniques [76,77]. I applied both to explore the creation of formal knowledge.

In this section I discussed the general characteristics of the methodology used in this thesis. In the next sections I will discuss the specific methods used for data collection and analysis in each study in more detail.

3.2 Description of methods: meta-narrative literature review on mindlines

Brief background

As mentioned in chapter 1, Gabbay and Le May published a prominent article in the British Medical Journal in which they challenged the 'over-rationalist model implicit in evidence-based health care' [10,84]. In an ethnographic study in UK general practice, they showed that clinicians only rarely accessed research findings, clinical guidelines and other types of formal knowledge directly. Rather, they preferred to rely on what they called 'mindlines', defined as 'collectively reinforced, internalised tacit guidelines, which were informed by brief reading, but mainly by their interactions with each other and with opinion leaders, patients, and pharmaceutical representatives and by other sources of largely tacit knowledge that built on their early training and their own and their colleagues' experience' [10].

Much subsequent research in mainstream health services research pursued and confirmed the negative finding of this study – that doctors rarely consult written guidelines when making clinical decisions. Studies sought to identify 'barriers' to guideline implementation on the assumption that more assiduous following of guidelines by individuals would lead to more evidence-based care – see for example [85–87]. A somewhat smaller literature (reviewed in this thesis) aligned with Gabbay and Le May's positive finding – that doctors follow mindlines – and sought to characterise, explore and occasionally critique the concept of collectively embodied tacit knowledge and how it links to the goal of evidence-based practice.

As quoted in chapter 2, the early EBM movement explicitly set out to 'de-emphasise intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient ground for clinical decision making', while stressing the instrumental role of evidence from research (especially the 'gold standard' design, the randomised controlled trial) in clinical decision-making [20]. Although Sackett et al. later softened this stance by writing that 'EBM is not restricted to randomised trials and meta-analyses' and should be seen as 'the best external evidence with which to answer our clinical questions' [3], at a philosophical level EBM seems to rest on a Cartesian view of knowledge as 'facts' that are stored in the heads of individual practitioners and/or in formal knowledge repositories, separate from the physical body, independently verifiable and distinct from values. Notwithstanding recent calls by the 'real EBM' campaign for a broadening of the parameters of EBM [1], such a conceptualisation would see mindlines (at best) as a lesser form of knowledge and (at worst) as not really existing.

Research question and study design

In this review I wanted to discuss whether and how mindlines have influenced, or should influence, the EBM movement. To explore the impact on EBM, I sought to document how the concept of mindlines was picked up and applied by researchers and practitioners following Gabbay and Le May's seminal paper. As mentioned above, I undertook an interpretive synthesis of this literature in collaboration with my supervisor Trisha Greenhalgh (TG), drawing on the principles of meta-narrative reviewing [12,80]. I also sought to further advance how mindlines have challenged EBM by drawing together findings from these papers as well as revisiting Gabbay and Le Mays' detailed explanation of mindlines in their 2011 book and bringing in relevant literature from a

wider range of disciplinary traditions, including philosophy of science. In this systematic review, we argued that to study knowledge that supports clinical decision making, conceptualised as mindlines to its full potential, we need to break out of the constraining notions of 'dissemination' and 'translation' (both of which imply a Cartesian view of knowledge) and focus more on the embodied nature of tacit knowledge and the interactive processes of knowledge creation.

Data Collection

I started by searching for any article with a reference to the word mindline or mindlines using the search query mindline* in the PubMed, Web of Science and OvidSP reference search engines. In PubMed I selected all NCBI (National Center for Biotechnology Information) databases up to 2014 and in Ovid the Embase 1996 to 2014 Week 50, HMIC (Health Management Information Consortium) database 1979 to November 2014, Journals@Ovid Full Text December 16, 2014 and Books@Ovid December 11, 2014 databases. I then searched and included articles that cited Gabbay and Le May's 2004 paper on Web of Science and screened Google scholar and books for any further relevant references. I merged the results and excluded hits in which the term 'Mindline' was an author's name.

Searching for mindline(s), we found one book, 139 articles in PubMed Central (PMC), 11 articles in PubMed, 14 results in Web of Science and 69 results in Ovid. When merged, 196 results remained. We found 213 results citing Gabbay and Le May's paper and another 24 publications by screening Google scholar and books. After merging these results, removal of duplicates and excluding 12 articles by or referring to authors named

Mindline, 340 references remained. TG checked 10% of the sample, before analysis. The complete reference list is in appendix 1.

There were no publications before 2004 in the remaining dataset; Gabbay and Le May appear to be the first researchers to use the word 'mindline' [1]. They subsequently elaborated on the concept in a book [12]. Although the paradigm-shifting potential of the mindlines concept was the subject of an early BMJ editorial [2], uptake of the term by the wider academic community was slow. A study from 2006 was unable to detect 'mindline' as a term in a large sample of articles from 12 major medical journals [13]. Most papers referred to Gabbay and Le May's work on mindlines in an implicit way (for example, in 195 publications, we were able to confirm that the term 'mindline(s)' was not used at all).

Data Analysis

In an initial familiarisation phase, we independently read the articles repeatedly and highlighted sections where the term 'mindline(s)' was used. We then collaboratively sorted the articles into paradigmatic groups (i.e. by the philosophical assumptions, theoretical models and methodological approaches shared by different groups of authors). In a subsequent synthesis phase, we compared and contrasted these different paradigmatic perspectives using narrative methods.

In explaining our findings, we adapted a taxonomy originally developed by Orlikowski and Iacono [88] to classify the different ways in which the term 'IT [information technology]' was used in papers describing 'IT research'. The reason for this was that

the notion of the 'nominal' view from this taxonomy applied very well to many papers in our sample (i.e. papers in which mindlines were mentioned by name but never defined or explored). To this, we added an 'in practice' view (i.e. observational studies of how mindlines are enacted in real-world clinical practice), a 'theoretical and philosophical' view (in which mindlines were included in a wider theory of knowledge sharing and/or collaborative practice or articles discussing the ontological and/or epistemological assumptions of the concept of mindlines); and a 'solution' view (i.e. proposals for how the development of evidence-based mindlines might be promoted and supported). Importantly, many articles could be incorporated into several of these views and these framings are not intended to be mutually exclusive or fixed beyond debate.

The results of this study are presented and discussed in Chapter 4.

3.3 Description of methods: theoretical work on purification, inference and the concept of truth

Central themes addressed in this thesis are not only about the nature of knowledge, but also "how does knowledge relate to reality, how do we reason with knowledge and what do we mean by truth?" as laid out in chapter 2. I wanted to consider these dimensions both from a philosophical perspective but also from a practical one, specifically in relation to guideline development. Accordingly, I used two complementary approaches in parallel, with each informing the other: a wide-ranging literature review and a naturalistic study of the real-world challenges faced by guideline developers in relation to deciding what is true and why. I describe the former of these linked approaches below, the latter in the following sections.

I developed theoretical work in the form of papers that were published throughout the DPhil project in collaboration with my supervisors as well as with GIN AID, a working group of the Guidelines International Network. This approach was aimed to capture literature and ideas from relevant research traditions, including epistemology and knowledge translation. I attended and helped to organise several seminars and workshops to present ideas and work in order for researchers and experts to contribute to, and comment on, the identification of key papers and literature in their field. Additional database searches and snowballing techniques were used to refine emerging meta-narratives and the overall synthesis.

Considering the way that knowledge relates to reality in EBM, I was inspired by an idea from my supervisor Eivind Engebretsen (EE) to examine Latour's book on modernity and the notion of purification of reality in EBM. I drafted the paper to which my supervisors commented which resulted in the paper as published in Journal of Evaluation of Clinical Practice in 2017 under the title "Has evidence-based medicine ever been modern? A Latour-inspired understanding of a changing EBM" [89].

To produce the theoretical work on issues on how guideline producer reason using knowledge in EBM, I met several times with other members of the GIN AID Knowledge Working Group to discuss the issues of using different knowledge and reasoning in guideline development, based on examples from their own experience and from published literature [90].

The AID Knowledge working group aims to contribute to bridging the gap in methods for Appraising and Including Different Knowledge. To support this ongoing challenge, the GIN AID Knowledge Working Group aims to identify methods and promising initiatives for appraising and including a range of knowledge sources in guidelines; facilitate sharing experiences with such methods and their application in practice through workshops at GIN conferences and elsewhere; identify areas / questions for methodological development; work towards reporting standards for knowledge appraisal and inclusion, and; develop a guide for appraising and including different knowledge [2,90].

I conceived the central argument of the purpose of guidelines and evading the problem of induction and drafted the first outline drawing on work by Ian Hacking. Through workshops, group discussions and iterative versions, the paper evolved in collaboration with most notably Teun Zuiderent and Beth Shaw, which resulted for instance that the notion of integration was added. As explained in the paper, the work was presented after several rounds of revision, in a workshop at GIN 2016 conference in Philadelphia. Input of workshop attendees was compiled, reviewed and included in the paper originally published in BMJ Evidence-based Medicine in 2018 “Different knowledge; different styles of reasoning. A challenge for guideline development”, a of which I present an adapted version in this thesis [90].

The theoretical work on the concept of truth was originally conceived by me to explore the notion of bias and the nature of reality in EBM. EE inspired me to read several publications from Isabelle Stengers in English and French on her notion of bias as inter-esse. Further suggestions, comments and adaptions from KH and TG resulted in the of “Rethinking Bias and Truth in Evidence-based Healthcare”[91], a paper originally published in Journal of Evaluation of Clinical Practice in 2018 of which I present an adapted version in this thesis.

The results of this theoretical work are presented in Chapter 5, 6 and 7.

3.4 Description of methods: a digital ethnography of mindlines online

Brief introduction

As mentioned in chapter 1, the mismatch between the knowledge captured in guidelines and the knowledge actually needed for clinical practice does not appear to have been fully anticipated by the pioneers of EBM. They argued that clinical expertise and patient preferences should be "integrated" with best research evidence [3]. Contrary to how some critics depicted EBM, "best" evidence was not considered to be synonymous with a simple and restrictive hierarchy of evidence, as some clinical questions are best addressed using study designs other than RCTs or because there are some questions for RCT evidence that is impossible to obtain or unavailable [4].

Despite this early call for a pluralist approach to evidence in guideline development, standards and checklists for assessing the quality of guidelines (notably the GRADE recommendations [5]) can sometimes depict an overly hierarchical approach, inadvertently privileging RCTs even when these are not appropriate and making it difficult to give appropriate weight to knowledge from non-RCT study designs.

In this discussion, it is important to note that EBM and EBHC have helped to develop numerous methods for the development and management of formal, explicit knowledge such as the structured population-intervention-comparison-outcome approach to formulating research questions, evidence hierarchies, search tools and strategies, techniques for statistical summation of trial results (meta-analysis), and so on. However, with few exceptions, EBM scholars have placed less emphasis on

investigating and describing how tacit, embodied, and practical knowledge is constructed in informal settings.

Research question and study design

The question 'how is knowledge actually developed and shared in informal communities of clinicians?' formed the focus of the first primary empirical research study presented in this thesis.

This study was published in a paper co-authored with my supervisors called "How Knowledge Is Constructed and Exchanged in Virtual Communities of Physicians: Qualitative Study of Mindlines Online"[92] in the Journal of Medical Internet Research. I conceived the study and its design, gained ethical approval and collected the data in the UK and the Netherlands. We were helped by research assistant Kristiane M. Hansson (KMH) from the University of Oslo in the collection of the Norwegian data. TG, EE and KH facilitated data collection and contributed to the analysis of the data and writing the research paper.

The design of the study was a digital ethnography of online professional communities, aimed to reveal how informal collective knowledge can be built and shared. We set out to study GPs reflections-on-practice and knowledge-for-practice. This is one step removed from actual practice, but important in its own right for several reasons. Although practice is not studied directly (hence, it is not possible to document how tacit, embodied knowledge influences specific instances of clinical practice), this methodology

was aimed to reveal for the first time how mindlines develop dynamically through multiple contributions to topic-based threads.

New technologies (specifically, electronic social networks, online bulletin boards, and email lists) offer new possibilities for collaborative knowledge creation that are not built into the traditional EBM framework. We sought to explore these possibilities by examining virtual networks of physicians to inform closer links between the development and use of clinical guidelines and the mindlines that emerge informally among communities of clinicians. This part of the primary research project focused on three virtual social networks of physicians in different countries. Our objectives were to describe the form and nature of knowledge, and the practices involved in knowing, of these practitioner communities; identify how mindlines develop in such communities; and explore how a broader set of knowledge sources influence (or why they fail to influence) the clinical community.

Gabbay and Le May's theoretical work on mindlines was derived from an extensive review of the philosophy of knowledge [15], as well as extensive ethnographic research comprising direct observation of clinical practice and face-to-face discussions (e.g., among local peers in practice meetings). This empirical work, undertaken at a time when clinicians' peer interactions increasingly occur virtually, asynchronously and in large online communities [33-35], sought to complement the original approach taken by Gabbay and Le May.

It has been shown that more formal online communities, including social media, have the potential to empower health care professionals and patients to apply knowledge by involving them in the intermediation and development of that knowledge [37,38]. Some virtual communities are already successfully used to create and share new knowledge using a strict structured format of questions and replies. For example, the network sermo.com that originated in the United States and claims to have 600,000 clinicians participating, lets its members post and answer multiple choice questions, with some additional features such as adding new answers to those questions to choose from and masking other member's answers for a few weeks [39].

Web-based networks are one form of informal community of exchange between GPs, but they are not unique. Informal communities of doctors in hospitals have been explored extensively, for instance, in the study by Haldar et al [36]. But web-based communities have some specific characteristics that make them particularly suitable for empirical study.

First, the dialogue is not limited in time and space (e.g., not limited to a particular medical institution where doctors meet). As such, they offer an efficient way of researching collective knowledge involving thousands of GPs from hundreds of sites and practices. Second, the access is less regulated than many other comparable arenas. There is some kind of selection (you must be a doctor, you must have sufficient electronical skills, etc), but the selection is less formal and less strict than in many other settings where doctors interact and exchange knowledge. Third, as private communities of clinicians, they provide an environment to freely discuss topics without the presence

of patients or representatives of the pharmaceutical industry. Fourth, there usually is limited formal censorship of the dialogue by for instance editors. The most important censorship is the governmentality or self-regulation of the group (e.g., members do not want to ask “stupid” questions).

We took virtual networks to be ‘knowledge artefacts’, not so much as carriers of objective knowledge, but as in the words of Cabitza et al. call “collaboratively created [inscribed artefacts], maintained and used to support knowledge-oriented social processes (among which knowledge creation and exploitation, collaborative problem solving and decision making) within or across communities of practice” [33,34]. As such, virtual networks and mindlines share the inherent duality of knowledge artefacts as holders or as catalysts of knowledge at the same time. An important property of knowledge artefacts is under-specification: they consist of an unstable, disjointed collection of shared representations, in the form of descriptions, images and expressions, which function nonetheless perfectly to support the knowledge activities of the members of the community. The extent of under-specification is subject to the effort it takes to maintain the knowledge artefact, the need to avoid information overload and the degrees of freedom to allow users to interpolate the representations in their specific context as well as communicate what cannot be expressed in words. When looking at knowledge artefacts, it can be hard to see how specific terms and symbols in themselves could have any practical purpose. However, their value lies in the knowledge that they invoke when they are interpreted by users [33].

Data collection

We obtained access to three social networks which provided contrasting but comparable datasets: a Facebook group of UK-based clinicians called 'Tiko's GP Group (TGG)', a virtual network set up by the professional GP societies in the Netherlands called 'HAWeb' and a network of physicians in Norway called 'Eyr'. The TGG group of UK GPs consisted of approximately 2900 members. The HAweb group consisted of more than 30,000 members and Eyr had 2300-2400 members. All networks have a dedicated administrator or administrative team moderating membership and online activities. TGG and HAWeb networks are 'closed', they can be accessed only after specific approval of a group administrator who confirms that the participant is a clinician (TGG) or a member of a medical society (HAWeb). Eyr is not closed, but accessible for anyone after registering; however, in practice almost all members are medical doctors.

I had already become a member of TGG and HAWeb years before the start of this DPhil project. I had helped to develop HAWeb for the Dutch GP Associations NHG and LHV in the Netherlands after seeing and signing up for similar network groups online. In the UK I heard from young colleagues about TGG on Facebook and signed up to get GP relevant updates and news. I and KMH signed up for Eyr specifically for this research project. We contacted the administrator of each network to explain the project and ask for consent. We did not participate in any group discussions except the one to highlight the nature of the research and give participants the opportunity for feedback.

Based on the large number of posts on these virtual networks and to allow comparison across networks, we chose to restrict the focus of our study to a single topic that was

viewed as a common clinical challenge, with current uncertainty and debate about clinical decisions, in all three countries. The issue we selected was raised levels of cholesterol (detected by a blood test and viewed as a risk factor for cardiovascular disease) and statins (cholesterol lowering drugs such as simvastatin, atorvastatin and related molecules). For years, statins have had a prominent position in debates in medical communities across many countries. Questions asked about statins include which sub-groups of people with raised cholesterol levels will gain a significant benefit from taking a statin [93] and how to deal with patients who cannot tolerate statins – for example because of the common side effect of muscle aches [94]. However, the prevailing debate in the medical literature concerns the interpretation of evidence on whether the benefits and harms of statin therapy have been over- or underestimated [95–97]. Some scholars argue that almost everyone should take a statin as it may prolong their life and side effects are rare; others that side effects are much commoner than the results of RCTs suggest and people with minimally raised cholesterol and no other risk factors would be better to avoid these drugs. As this debate is heated and still ongoing, statins form a rich and interesting subject to explore mindlines online.

Using the native search engines in these networks, we aimed to look for online posts from 2013 to 2015 that contained (parts of) the following words: statins/statin, cholesterol, cholesterol-lowering treatment, cholesterol-lowering drugs, hypercholesterolemia, familial hypercholesterolemia, HMG-CoA reductase inhibitors and their Norwegian and Dutch equivalents.

Ethical approval

Obtaining ethical approval for this online research proved to be challenging. There were several ethical considerations and risks to participants.

- Members needed to know that they were part of a study and should have the opportunity to opt out and give feedback.
- As we had chosen to study closed groups, participants were expected to discuss matters in a peer-to-peer fashion. Our research could be considered an injunction to this by administrators and members as publication could be seen as exposing intimate group discussions that were not intended for non-peers.
- We considered the risk of accidentally exposing patient data. However, the group consisted of GPs who are bound by strict professional rules of confidentiality. Statin use is very common, so identification of patients was seen as unlikely. When any identifying information such as photos are posted this is with the patients' consent. But my experience as a member of the group had already shown that cases were discussed only in a very general way.
- The Facebook group administrator had put extra policies in place and monitors posts. Identifiable patient information was therefore very unlikely to be posted. Links to other identifiable persons or contributions of persons outside of the group appeared to be limited, but if these were to be encountered, they would not be used.
- We also contemplated that participants may disclose information of a personal nature such as their own experience with a disease. In this way participants

were not only clinicians discussing anonymous patients, but potentially identifiable patients themselves. We would have to be careful to use this data to preserve anonymity of the participants.

- Special consideration was also needed for any verbatim quotes as the search function of Facebook makes it possible for other group members to identify the original contributor.

In order to mitigate these risks, I discussed our approach with social networks researchers Farida Vis and Natasha Whiteman of Manchester University. They provided ethical background information on virtual social network studies and also helped to identify ethical considerations of this study as mentioned above [98,99]. We discussed these concerns and how to deal with them with the ethical committees in each country.

I planned to seek consent from the social network administrators first and then present the research plan in an online discussion post in the networks, asking members to discuss whether this study would be acceptable to perform. I arranged to explain any potential risks of the study, its value for research and ask for cooperation to the online members in a post and answer further questions and feedback. I also planned to pre-published some of the findings to the group to check any results as well as the final paper. In collaboration with TG, I considered that the community could raise concerns in their comments, which would inform investigators and group administrator about which precautions need to be taken. In case of serious concerns, we recognised that this could lead to a major change in methodology or in an extreme case, the cancelation of this part of the study.

We first thought of using verbatim quotes for publication. But to maintain anonymity of participants, we decided to translate quotes from Dutch and Norwegian and paraphrased posts from English to avoid connecting them to members using the virtual networks' native search engines. We still asked individual participants if they approved of publishing the paraphrased quotes and whether publishing these quotes would have any unintended risks for the individual and for the group now or in the future.

The research was approved for each network in each country separately. By the University Research Ethics Committee (QMREC) at the Queen Mary University London in the UK under reference number 2014/82, by NSD at the University of Oslo in Norway under reference number 48032 and by Research Ethics Committee (REC) at the Radboud University Nijmegen Medical Centre in the Netherlands under reference number 2016-2680.

The administrators of each network also approved. As planned, I asked members of the online networks first if they approved of the study using posts explaining the background and potential risks. This resulted in a debate in the Facebook group. Most participants approved of the study, but one member considered the study an intrusion to a community. I agreed not to use postings from this individual in the research. Other participants did not agree with his point of view. I discussed this debate with the QMERC at Queen Mary University, who took note but approved continuation of the study.

Data analysis

Using above mentioned collection methods, more than 1400 relevant posts over the three platforms were found. Since our analysis was focused on the generation and application of mindlines in online conversations between doctors, we needed an analytical approach that reached beyond mere thematic descriptions of these posts. We therefore used an iterative part thematic descriptive and part interpretative approach [100] to analyse and synthesise the data with mindlines as a reference model. Firstly, we were interested in themes and topics related to the participants' use of knowledge and knowledge sources, such as references to guidelines, research-based evidence, personal experiences, reported patient experiences with statins, and so on. Secondly, we wanted to capture the participants' "acts of knowing" in terms of the logics underpinning their reasoning and the interactions in the online communities. We were interested not only in what sources they built their arguments on but how they built and exchanged their arguments. What characterized their "styles of reasoning" [79,101]? How did they address and invite each other into discussions? How were patient problems articulated and what sorts of responses were given by whom?

The empirical data from online discussions were subject to preliminary analysis by each researcher independently with the aim of identifying and classifying both the above-mentioned aspects of the knowledge creation in the conversations. Secondly, we discussed our findings as a group and compared and weighed our different descriptions against each other and against Gabbay and Le May's description of mindlines [10]. A draft of the descriptive research results was fed back to the online communities for

review and to request input for topics for discussion. Finally, in an interpretive synthesis process, the topics discussed by the online communities, key findings and limitations of the study were drawn out through reflection and discussion among team members. The results of this study are presented in Chapter 8.

3.5 Description of methods: ethnographic study of guideline development panels

Brief introduction

Since mindlines play such an important role and provide an alternative view on clinical knowledge creation, they could potentially inform the development of guidelines that clinicians will follow as they are meaningful and useful for everyday practice. But this kind of 'soft' knowledge is not easily accessible to guideline development groups – nor is it easy to verify or validate. How and when should we supplement evidence from RCTs with 'softer' evidence? To what extent, and how, can we predict outcomes for a single patient? Who should decide what counts as 'valid' knowledge and why [16]? In this second primary research project, I set out to look into formal evidence-based knowledge creation by professional institutions, in an effort to uncover tensions between guidelines and mindlines that indicate the limits of the EBM paradigm.

In order to do this, this research study contemplates what Tonelli and Upshur have called "epistemic uncertainty", which they define as "[t]he challenge of applying clinical research results to individual patients; dealing with multiple sources of (at times conflicting) information, guidelines, and principles; and determining what level of confidence to have and convey in a medical conclusion" [102]. Whereas literature on epistemic uncertainty usually addresses how it emerges and how students could be taught to deal with it in the clinical encounter [103–105], in this research study I shift the analytic lens to study efforts to reduce epistemic uncertainty in the guideline development process.

Tonelli and Upshur exhort medical educationists to teach students not merely to accept epistemic uncertainty in clinical practice but to actively explicate and mitigate that uncertainty [102]. Eivind Engebretsen e.a. have argued that far from being a negative influence on clinical practice, epistemic uncertainty can be a positive force, prompting the clinical imagination through reflexive questioning and using active awareness of multiple possibilities to support lateral thinking in the diagnostic process [104].

As noted in chapter 2, clinical practice guidelines have been depicted as a "central instrument of EBM" [18], aiming to reduce epistemic uncertainty by making diagnosis and treatment more data-driven. A small but important tradition of research into the guideline development process has shown that distilling guidelines from research-derived knowledge is not unproblematic. Moreira's meticulous ethnographic studies of local guideline development panels in the UK, for example, revealed that the process involved four interacting repertoires of collective judgement: 'science' (in which panel members discussed the technical robustness of research evidence); 'practice' (in which they addressed usability); 'politics' (in which they considered the guideline's likely acceptability to different stakeholder groups) and '[due] process' (in which they reflected on the adequacy of the guideline development process itself, including quality of group interactions and whether conflicts of interest had been fully declared) [106].

This and other studies (see for example [107–110]) have used a sociology of scientific knowledge lens to show that in distilling the "truth" from research studies, guideline development has long been shown to be a contact sport in which science abuts politics and practical reality.

In this study, the main lens was not sociological but philosophical, focused on the ontological and epistemological basis of guideline development. Accordingly, whilst I acknowledge Moreira's political, practice and process repertoires, the main empirical focus was not merely to describe what is happening but to understand whether and how these repertoires could help to ultimately reduce epistemic uncertainty in clinical practice.

Research question and study design

For this project I collaborated with TG, EE and KH. I conceived of the study and managed data collection; they helped to finalise the study design, gain ethical approval and analyse the data. The research questions were:

1. How do guideline development panels combine (or why do they find that they are unable to combine) different kinds of evidence such as non-RCT research designs, routinely collected data, patient experience data and clinical experience and wisdom?
2. How is the work of guideline development panels influenced by different perspectives on epistemic hierarchies and approaches to uncertainty?
3. What are the implications for teaching students about epistemic uncertainty in clinical practice?

We set up a conventional ethnography of guideline development panels to explore how broader knowledge sources including mindlines influence the development of clinical

guidelines and to understand formal medical knowledge guideline development processes, the behaviour, argumentation and views of people involved in those process, their roles and beliefs regarding these activities as well as adaptation to new processes and techniques affecting guideline development.

We sought to include panels involved in guideline development and documents they handle and produce. The method of research had participatory as well as observational components.

We used purposive sampling to select guideline institutions and panels [111]. In order to gain in-depth data of EBM in practice as theoretically intended, we wanted to select panels on a variety of topics at guideline developing institutions which are regarded as using rigorous and high standards of EBM methods by the guideline developing community. We also wanted to include guideline panel meetings in several countries, since processes for guideline development and the recommendations they produce on the same clinical entity vary considerably internationally, seemingly influenced by cultural factors, including habits, patient expectations and the healthcare system [112–114].

Through our international network and connections with members of the guideline international (GIN) network, we gained approval to study panels at the National Institute for Health and Care Excellence (NICE) in the UK, the Dutch College of General Practitioners (NHG) in Netherlands and the Norwegian Directorate of Health. NICE

produces and publishes national clinical guidelines as well as guidance on the use of health technologies, social care and public health. The Norwegian Directorate of Health produces guidance on public health. NHG produces all guidelines in primary in the Netherlands and provides significant input on most guidance in secondary care and public health. At NICE we were able to conduct both observational as well as participatory research as I was already a member of a guideline panel at NICE. At NHG and the Norwegian Directorate of Health we conducted observational ethnography only.

Purposive sampling was used to organise semi-structured interviews (via Skype, telephone or in-person) with members from the guideline panels, technical committees and management from guideline developing institutions. Their recruitment was an on-going process occurring throughout the study via informal contact or formal nomination of the appropriate person by the guideline-developing organisation.

Data collection

Data was collected from meetings and documents from guideline panels as well as interviews with their members in a broad range of topics relevant to primary and secondary care, public health and social care.

I collected data both as a participant in my role as a standing member of a clinical guidelines update committee, as well as an observer of other guideline panels. Senior researcher David Wright (DW) further helped to collect data by observing several meetings at NICE. In Norway, Gina Fraas Henrichesen (GFH), a senior researcher from

the Knowledge in Translation (KNOWIT) research group, helped to collect that data in meetings at the Norwegian Directorate of Health I collected data at the Netherlands NHG.

Field notes were written to record the main discussions, questions, answers, narratives, impressions, explanations, and alternative interpretations. We did not take audio or video recordings of panel meetings. The analysis of documents focussed on publicly and non-publicly available documents produced by the guideline developing institute, committee and group members. At NICE this included the 'leading evidence to recommendations' (LETR) tables which listed the contemplations of the guideline panels when they make a recommendation.

Ethical approval

We first asked guideline panel members to discuss whether this study would be acceptable to perform in collaboration with the guideline developing institute's management or the online group's admin. We explained in writing to the organiser of the panel, the chair or in person before the panel the design and purpose the study, its value for research and asked the group as a whole as well as individual members for consent. We answered further questions and feedback beforehand, during and after meetings. We discussed our findings with a guideline development coordinator at each institution to get feedback on results, analysis and conclusions as well as to make sure our publications upheld participant anonymity as required by each of the institutions.

To maintain anonymity of all participants and interviewees procedures were followed to mask conditions, interventions and outcomes related to the guideline topic as much as possible.

We asked all panel members to fill out a consent form. Consent was also explicitly obtained from online members whose verbatim quotes we intend to publish in peer-reviewed papers. Additional official consent procedures at the guideline developing organisations were followed too. We discussed with the guideline developing organisations whether publishing our findings could have any unintended risks for the individual, the panels and the organisation now or in the future.

We obtained ethical approval for this research from the Social Sciences and Humanities Inter-divisional Research Ethics Committee (IDREC) at the University of Oxford in accordance with the procedures laid down by the University for ethical approval of all research involving human participants (reference number EQ C1A 16-28). In Norway we obtained ethical approval from the Data Protection Official for Research via the University of Oslo (project number 48032). In the Netherlands, the research ethics committee of the Radboud University Nijmegen Medical Centre declared that for this the study (File number CMO: 2016-2680) could be carried out (in the Netherlands) without an approval by an accredited research ethics committee and without explicit informed consent of the participants as it did not fall within the remit of the Medical Research Involving Human Subjects Act (WMO).

Data Analysis

Data collection ran from September 2016 to February 2019. We used purposive sampling [111] to select national guideline development groups in UK (National Institute for Health and Care Excellence [NICE]), Netherlands (Dutch General Practitioners Association [NHG]) and Norway (Norwegian Directorate of Health) – three countries with different healthcare cultures and contrasting processes for guideline development [112–114]. All groups were regarded by the guideline development community as applying high standards of EBM in their work.

With help of GF and DW, I observed 19 panel meetings (including 13 in observer role and 6 in participant-observer role as a member of the panel). We also undertook semi-structured interviews (remotely or face-to-face) with a maximum-variety sample of 10 participants recruited iteratively as the study unfolded and collected relevant documents. Details of data sources and their contribution are given in table 2. To disturb the process as little as possible, panel meetings were not audio-recorded, but we made detailed contemporaneous field notes with selected near-verbatim quotes.

Table 2: Overview of data structure

Approach	Description of sample	Type of data	Descriptive insights	Higher-order insights
Ethnographic observation of guideline development panel meetings	19 meetings lasting 3-9h each (120h total) - 10 in UK (NICE) - 6 in Netherlands (NHG) - 3 in Norway (Norwegian Directorate of Health)	Field notes focusing on actions and interactions of people involved in the guideline development process, including arguments (e.g. explanations and counter-explanations offered)	How panel members collectively drew on different kinds of research evidence and other documentary sources in making (and justifying) recommendations. Nature of arguments among panel members.	Insights into the process of "purification" (selection, abstraction and reconfiguration of research evidence to generate carefully shaped evidence statements targeting particular renderings of the patient e.g. as a member of a risk group [110]). Competing discourses and world views among panel members.
Semi-structured interviews	10 in total - 6 in UK - 2 in Netherlands - 2 in Norway	Interviews with guideline development group members, technical committee members, and managers from participating organisations	Individuals' views on particular topics, their assessment of the evidence, and their reflections on how the guideline development process had unfolded for those topics.	Addition of nuance to ethnographic findings.
Documents	> 200 documents (publicly available and not publicly available) - 136 from NICE - 64 from NHG - 10 from Norwegian Directorate of Health	1500 pages of agendas, minutes, draft guidelines, review protocols, presentations on review techniques, overviews of evidence, GRADE tables. In NICE, 'linking evidence to recommendations' (LETR) tables and chapters with detailed contemplations of panels	Material and visual properties of guidelines and interim documents. Elucidation of the (sometimes complex) chains of reasoning through which panels linked particular evidence sources to particular recommendations.	Insights into the guideline as a text that 'does work' (e.g. inscribing and stabilising particular renderings of patient, illness or risk state, and intervention).

In collaboration with my supervisors we used an interpretive approach to analyse data. Firstly, field notes, document excerpts and interview transcripts were thematically coded to identify first-order data (broad categories, themes, patterns and relationships through constant comparison [115]). To generate the higher-order data in the last column of table 2 above, we selected longer excerpts for more detailed discussion among all authors, inspired by Mol's philosophical work on studying multiple enacted realities [116]. For this, we focused mainly on what individuals and panels *did*, and what was brought into being through this action. Specifically, we considered how different

practices of guideline development enacted different understandings of EBM (while also, sometimes, contributing to a common understanding).

Finally, we summarized findings and interpretations through reflection and multidisciplinary discussion with guideline developers and panel members. To maintain anonymity of all participants and interviewees, we followed a validated process to systematically fictionalize conditions, interventions and outcomes related to the guideline topic [117].

The results of this study are presented in Chapter 9.

3.6 Conclusion

In this doctoral work, I sought to illuminate the tensions between guidelines and mindlines as indicators of how EBM reaches its limits as a Kuhnian paradigm, using qualitative and naturalistic methods. I have pursued three overlapping studies, beginning with an in-depth literature review on the concept of mindlines. I also undertook a digital ethnography to explore the creation of informal knowledge in groups of clinicians. Finally, I explored the development of formal knowledge using conventional ethnography.

Overall, my methodological approach allowed me to build a rich picture of the tensions in EBM guideline development in a way that could complement and extend more conventional approaches to the study of EBM/EBHC such as quantitative research of its adoption and achievements or purely theoretical critical analysis of its methods. Using a combination of methods as such was not so much about triangulation to define a single unified reality but about becoming what has been called an "insider and outsider at the same time", in order to be able to take and negotiate different standpoints simultaneously (an interpretivist position) [78,118].

The approach taken raised both practical and ethical challenges which I have set out in detail above. In sum, the main challenges were, to gain ethical approval and access to the online platforms and guideline producing institutions. In the next chapters, I will present the key findings of my empirical work, starting with the development of the dimensions of the theoretical framework as explained in Chapter 2. In the discussion in

chapter 10, I will return to consider the strengths and limitations of my methodological approach and the extent to which it allowed me to answer my overall research question and the various sub-questions. As noted in sections throughout this chapter, my supervisors have made various contributions to co-authored papers and have affirmed the nature of this input.

Chapter 4: On the concept of knowledge and knowledge creation: mindlines

This chapter is an updated version of a paper co-authored by my supervisor Trish Greenhalgh (TG) called '10 Years of Mindlines: a Systematic Review and Commentary', originally published in Implementation Science in 2015 [16].

4.1 Introduction

Building on the methods described in section 3.2, in this chapter I want to discuss my theoretical findings on *whether* and *how* mindlines have influenced, or should influence, the EBM movement. To summarise, in collaboration with TG, I sought to document how the concept of mindlines was picked up and applied by researchers and practitioners following Gabbay and Le May's seminal paper. We also aimed to further advance how mindlines have challenged EBM by drawing together findings from these papers as well as revisiting Gabbay and Le Mays' detailed explanation of mindlines in their 2011 book and bringing in relevant literature from a wider range of disciplinary traditions, including philosophy of science.

4.2 Results

Literature on mindlines that appeared after Gabbay and Le May's article could broadly be categorised into four groups representing different (but overlapping) paradigmatic

perspectives. First, there was the 'nominal' framing, in which authors referred (implicitly or explicitly) to mindlines but did not further explain or expand. We found 133 papers that would primarily fit this view. For instance, mindlines are mentioned in this paper as an unelaborated example of sub-optimal practice when the president of the Australian Patient Safety Foundation is quoted:

'There's a tendency to criticize evidence in order to maintain the status quo, […] medical practice is currently dictated by traditional approaches and 'collective mindlines'.' [119].

In seven papers, comments appeared to come from the guideline development community and most of them best fitted the nominal view. For example, a paper whose first author's affiliation is the Italian Cochrane centre, mindlines are referred to as 'anachronistic' [120] Another paper from the Canadian Thoracic Society (erroneously in our view) conflates mindlines with the rapid exchange of 'easily understood' information among practitioners:

'The literature suggests that health care providers seldom consult guidelines in practice; instead, they rapidly glean pieces of information from documents or colleagues with whom they construct 'mindlines' that inform clinical decisions. To emulate this practical knowledge exchange medium, guidelines should include information that can be easily understood and transmitted' [121].

We believe this recommendation reflects a naïve conceptualisation of the ontology of knowledge (seen as facts to be transmitted, simple ones being more transmissible than

complicated ones). As we argue below, to understand mindlines requires that we go beyond such a conceptualisation.

The second framing of mindlines we gleaned from the literature was the 'in practice' view, in which the term was used to explain the empirical finding that clinicians are rarely observed to follow written guidelines (but appear to follow mindlines instead). We found 76 papers that primarily referred to mindlines from this point of view. In these papers, the term 'mindlines' often seemed to mean 'consulting colleagues'. For example, one study showed that, like GPs, general surgeons most often turned to colleagues before using other sources of knowledge such as the internet, educational meetings or the library [122].

In a few studies, the 'in practice' view of mindlines closely reflected Gabbay and Le May's original theorisation. In a study on the role of tacit knowledge in how public healthcare groups planned initiatives, for example, the authors found that *'study participants used collectively reinforced tacit guidelines based on experiences and interactions in fluid communities of practice rather than drawing on research findings or explicit practice guidelines'* [123]. In a study on the influences on prescribing in general practice, Grant et al. found that practitioners used 'prescribing mindlines'. *'These were personal formularies developed from and informed by their experience of medication (including patient's experiences), specialist advice, discussions with their practice pharmacist and GP colleagues, and the practice's macro prescribing policy (if present). GPs rarely looked up information about medicines and relied on these prescribing mindlines. […] GPs relied on personal experience and social networks to update their mindlines'*[124].

In a cross-national study on what physicians gain and lose with clinical experience with diabetes Elstad et al write *'In tune with Gabbay & le May's 'mindlines' […], we found that physicians gained their experience in part through their professional interactions'*. They then quote an experienced German clinician who comments on comparing intuitive concerns: *'Happily we are two physicians working here… I say, 'It's funny, something is wrong with him'. And my colleague says, 'You know, you are right. Something is wrong. He simulated, or lied or something is rotten'. We nearly always agree. You cannot have that in an office where you work alone, but it's really good. We swap ideas on the patients'* [125].

Gabbay's and le May's ethnographic methodology is replicated in one of the few studies in the literature to tease out *what the key mindlines were* in a particular topic area [126]. In this qualitative study, the authors invoke three 'mindlines' of clinicians which appear to explain why malaria is over-diagnosed in Tanzania. In this example, however, one could claim that mindlines are merely depicted as unspoken rules of thumb or maxims that generally over-ride more specific and detailed formal recommendations:

'Rather than following national guidelines for the diagnosis of febrile illness, clinician behaviour appeared to follow 'mindlines': shared rationales constructed from these different spheres of influence. Three mindlines were identified in this setting: malaria is easier to diagnose than alternative diseases; malaria is a more acceptable diagnosis; and missing malaria is indefensible. These mindlines were apparent during the training stages as well as throughout clinical careers.' [21]

Chandler et al's study is cited in 33 papers in our sample, but most of these do not actually refer to the concept of mindlines (they cite the paper to support the statement that there is over-diagnosis of malaria).

Third, mindlines were sometimes framed as a theoretical or philosophical concept in publications (with or without an empirical component) whose main purpose appeared to be the development of theory. We classified 57 papers as predominantly this framing.

Many theoretical papers discussed mindlines in the light of Lave and Wenger's theory of communities of practice. For instance, Li et al combine the concept of mindlines with communities of practice and Nonaka and Takeuchi's knowledge creation cycle (which considers how formal codified knowledge is made tacit and disseminated among organisational members through observation and discussion, thereby becoming meaningful and applicable in practice) [127]. They state:

'Explicit knowledge is codified information such as peer-reviewed articles, rules, and guidelines, which can be readily shared among people. However, to apply this knowledge in practice, practitioners must make sense of the concrete information in the context in which it is used. This process of establishing meaning can be facilitated by discussions with colleagues and mentors or by observing how others apply the knowledge and then try it themselves.' [127]

Similarly, Crites et al. talk about the knowledge creation cycle and link this to the literature on learning organisations as described by Argyris and Schon [128]; they take from mindlines that external knowledge needs to be *'validated through informal team discourse and modified for practical application'* by working teams [129]. Ranmuthugala includes Gabbay and Le May's ethnography in an elaborate systematic literature review on why and how communities of practice are established in healthcare [130]. Soubhi et al offer a theorisation of the communities of practice literature, enhanced by the notion of mindlines, as it might be applied to multidisciplinary care for complex multi-morbidity [131].

Several studies see mindlines as empirical support for the tenets of social network theory primarily by taking the notion from Gabbay and Le May's work that clinicians rely on their peers to acquire knowledge. They set out to explore what personal relationships between clinicians exist - for example, to study the propensity towards EBM of physicians in relation to their 'coreness' (that is, closeness to the centre) in their social network [132]; to explore how the structure of patient-sharing relationships among physicians [133] is related to care patterns of high or low costs within hospitals [134]; and what factors affect the influence of certain physicians in a network on the thinking and practice of other physicians [135]. An alternative argument is that the concept of social networks neither confirms not refutes the kinds of knowledge exchange that are implicit in the concept of mindlines.

Other studies link mindlines to social influence. Lomas, for example, provides a useful overview of the literature on the cultural gap between research and policymaking – and

the essentially *social* (not technical) nature of successful efforts to bridge this gap [136]. He uses the concept of mindlines to underline the social nature of collective influence in both the research and policymaking communities. Nine publications in our sample explicitly criticised the political, economic and ethical dimensions of how research knowledge is generated (for example the tainting of research funding through industry conflicts of interest). However they usually did not link these critiques to mindlines directly.

In an article on clinical decision-making in Ghana, the authors used a novel theoretical framework of guidelines (explicit), clinicians' mindlines (tacit), and 'clientlines'. The last are a patient version of mindlines, consisting of *'client influences related to the preferences and pressures of the client and the wider family and community, including social, religious and cultural values and beliefs'* [137].

Papers that addressed the ontology (what is it) and epistemiology (how might we study it) of knowledge considered topics such as the difference between explicit and tacit knowledge, the nature of clinical reasoning and the validity and transferability of medical knowledge. For example, several authors referred to mindlines as a form of system 1 (intuitive) as opposed to system 2 (analytical) thinking in articles on clinical information processing and cognitive errors [138–140]. Bate et al explain:

'Dual process theory states that humans process information in two ways, termed System 1 and System 2. System 1 processing is an 'intuitive, automatic, fast, frugal and effortless' process, involving the construction of mental maps and patterns, shortcuts

and rules of thumb (heuristics), and 'mindlines' (collectively reinforced, internalized tacit guidelines). These are developed through experience and repetition, usually based on undergraduate teaching, brief written summaries, seeing what other people do, talking to local colleagues and personal experience. System 2 processing involves a careful, rational analysis and evaluation of the available information. This is effortful and time consuming. Data from a variety of environments demonstrates that human beings prefer to use System 1 processing whenever possible.' [140]

In a number of papers, however, this deeper understanding of mindlines was missing and authors simply equated mindlines to heuristics.

Walach et al propose a 'circular' model for linking evidence and clinical decision-making [63,141]. Like Henry et al. [64], they reject the hierarchy of evidence (randomised trials at the top, 'anecdotes' at the bottom) as valid only in relation to simple decisions about the efficacy of drug therapies. Most clinical decisions, they argue, are complex; they involve ethical and human decisions as well as scientific ones, and hence require the integration of multiple considerations and forms of evidence. The tacit knowledge of mindlines may be more appropriate for the complex and organic nature of real-world medicine than the 'if…then' structure of guidelines and decision support tools.

In a systematic review of knowledge exchange mechanisms, Contandriopoulos and colleagues consider knowledge in two essential forms: individual, that is, held in people's heads and translated (or not) into action by human will and agency, and

collective, that is, socially shared and organizationally embedded – a form akin to Gabbay and Le May's mindlines [142].

A final framing of the 'mindlines' concept, sometimes offered as a conclusion in an empirical or theoretical paper, was solution-focused: some authors considered the question of how to actively promote and support the development of valid embodied/collective knowledge or 'evidence based mindlines'. In 28 sources this seemed the main purpose of the article. For example, the challenge was expressed (somewhat obliquely) in a letter by Glasziou, in which he recognises the concept of mindlines in clinical practice, but is worried that they could *'supply counterfeit evidence. […] a puzzle remains: how do we get valid memes into the mindlines while not driving out the wisdom of experience?'* [143].

Two studies describe efforts to set up (and influence the behaviour of) communities of practice among researchers [144] and doctors [145] by employing a facilitator and co-ordinator respectively, though each of these studies mentions mindlines only in passing. In a more theoretically-informed paper, Soubhi et al's model of communities of practice in multimorbidity care is also solution-focused, emphasising relationship-building and collaborative learning as the basis for developing mindlines. They hope that qualitative and quantitative research could *'examine how primary care physicians develop mindlines and how they test them to eliminate harmful ones and standardize others into routine practice'*.

Others call for knowledge brokers [136], transformational leaders [146] thought leaders [147] and individuals more generally [144] to alter, expand and embed new knowledge through social influence within (and indeed extending beyond) existing communities of practice. These proposals echo and flesh out Gabbay and Le May's original exhortation to make sure that the knowledge circulating within communities is based on sound research [10]. In their systematic review, Contandriopoulos et al suggest that knowledge may become collectivized through a variety of mechanisms, including efforts to make it relevant (timely, salient, actionable), legitimate (credible, authoritative, reasonable), and accessible (available, understandable, assimilable) and to take account of the assumptions and priorities of a particular audience [34].

Reeve et al. draw parallels between mindlines and the (subtle and often overlooked) skills of expert generalism in GP practice [148]. These authors offer a four-phase approach to developing generalist expertise: sensemaking (popularising the concept of generalist expertise and raising awareness of it across a community of practitioners), engagement (influencing practitioners to prioritise this issue), action (e.g. delivering education, promoting scholarship as part of professional practice) and monitoring (measuring the impact of this approach).

In an observational study to assess how 'evidence based' GP consultations were, Zwolsman et al. observed that GPs were often unable to account for the source of their knowledge (personal experience or research evidence) and that rapid, intuitive decisions, suggesting a predominantly tacit form of knowledge, were the norm [149]. The authors felt that making all steps in the decision-making chain more explicit (for

example, justifying the chain of reasoning to the patient) would surface gaps in knowledge, inform further learning and – in the long run – make practice more evidence based.

In contrast, Levine et al. suggest the opposite. They propose making guidelines more *implicit* by transforming them in to aphorisms which they define as *'succinct sayings that offer advice'*. These should function as *'interface between intuitive approaches to make rapid decisions, and the implementation of specialty-specific clinical guidelines'* [150].

4.3 Discussion

Our systematic narrative review revealed a relatively sparse literature on the important concept of mindlines, first introduced in 2004 by Gabbay and Le May. Whilst we included 340 publications in our final dataset, some authors appeared to have (with the best of intentions) used this term incorrectly and naively. But others had recognised and explored mindlines' fundamental philosophical challenge to EBM. The relatively limited contributions from the guideline development community suggests that the concept has been largely ignored by the guideline industry – perhaps because, as American physicist and philosopher of science Thomas Kuhn wrote, *'the proponents of competing paradigms practice their trades in different worlds'* [14].

'Knowledge translation', one-way transfer of knowledge from producer (research) to user (clinician) [151,152], remains a (contested) policy challenge. Multiple research

traditions have contributed to a vast body of literature on how evidence from research is disseminated [153,154]. Underlying this literature is a fundamental hope for the possibility of optimising the 'intermediation' of knowledge – that is, the managed processes by which practitioners can be supported to interact with knowledge [155]. But an alternative metaphor that goes beyond the concept of 'spreading' good ideas is needed [156].

In their book from 2011, Gabbay and Le May explain in great detail the origins of mindlines, their implications and related theories [15]. Mindlines fit a view that knowledge is not so much a set of external facts that are waiting to be 'translated' or 'disseminated' but a more fluid and multi-directional phenomenon in which knowledge is 're-created' in different contexts by different people again and again as previously postulated by Mol and Nonaka [116,157]. From this perspective, improving knowledge intermediation is more like maximizing the opportunity to *create* knowledge. How this might be achieved differs from the EBM paradigm in a number of ways uncovered by Gabbay and Le May throughout their book. We explore these philosophical questions further below.

Reality – single or multiple?

Firstly, mindlines question our assumptions about reality. Without doing full justice to the vast literature and on-going debate on the philosophy of knowledge, we acknowledge in particular the philosophical difference between naïve realism (in which there is one reality, which exists independently of human thought and can ultimately be

known by everyone) and more pluralist philosophies (such as subjective idealism and critical realism) which assume multiple realities, none of which are fully shared [158].

The possibility of multiple realities explains why different national guidelines for even simple conditions like urinary tract infections draw different conclusions from the latest research [114]. Naïve rationalists (some protagonists of EBM, for example) think that we will overcome these differences and ultimately have a clear set of 'facts', recommendations, international clinical guidelines and policies that everyone except the misguided and ignorant will all agree upon. In contrast, as Gabbay and Le May point out [15], the concept of mindlines presents us with the idea of a shared (but by no means homogeneous) reality consisting of multiple very individual and temporary realities of people: clinicians, researchers, guideline makers and patients. Mindlines offer a view that the number of guidelines on any topic will never cease to expand, because we expect new individual realities and scientific paradigms to emerge continuously.

The nature of knowledge

Absent in the conventional EBM paradigm, but very present in the concept of mindlines, is the acknowledgement that, to use Polanyi's phrase, *'we know more than we can tell'* – that is, not all knowledge is conscious or explicit [159]. Gabbay and Le May discuss how mindlines include 'knowledge-in-practice-in-context'; practical knowledge formed not only by the setting but also the need for that knowledge [15]. Conventional EBM views knowledge narrowly as factual data, only a tiny fraction of which are relevant to the decision at hand (and that it will be largely self-evident which 'facts' are needed). The mindlines concept envisions a 'negotiating space' [15] where clinical decision-

making by clinicians and patients involves a process of reduction and prioritisation from a vast realm of potentially relevant knowledge of different kinds.

In this sense we would argue that mindlines stress both the act of, and the need for, reducing possible options for action. EBM does not reject the idea of reducing knowledge – for example the guideline development process encourages the dismissal of evidence and knowledge that is considered of low quality [2,160]. But in mindlines, the process of reduction applies not only to explicit knowledge. All kinds of knowledge are competing for attention [15,142]. In clinical consultations, not everything is or can be taken into account; there is limited time [161], our brains do not process everything [140], we are forgetful and the 'whole' story is not told to us. Reduction is an essential and ever-present process to create knowledge and experience reality. As Heisenberg wrote *'what we observe is not nature in itself but nature exposed to our method of questioning'* [162]. Arguably, what we observe as clinicians is not reality itself, but reality exposed to our method of reducing or filtering the various potentially relevant streams of knowledge of which we are consciously or unconsciously aware and from those, constructing a picture of current reality.

As we explore a clinical case by questioning, examining and testing, some things get more 'certain', but the overall picture will not necessarily become clearer. For example, in the case of abdominal pain, a stool sample makes us more certain about the organisms involved. However, it says little about the social context in which a particular type of pain occurs and recurs. If we ask a patient about her daily life, we may build a fuller, clearer and more holistic understanding of the abdominal pain – and the stool

sample may become less of a priority. Indeed, an incidental growth of a pathogen in the stool sample May confuse and distract us if the 'real' cause of the abdominal pain is migraine, marital stress or the side effects of medication. Reducing in the spirit of mindlines is a creative process, not a reductionist one, like carving a particular 'abdominal pain' out of a piece of buzzing, blooming reality [163] with an infinite number of dimensions.

How the 'truth' is arrived at

Gabbay and Le May offer many detailed examples of how clinicians co-construct knowledge and discuss several theories that help to understand these processes such as the knowledge creation cycle developed by Nonaka and Takeuchi explaining the growth of tacit knowledge in organisations [15]. But unlike the papers we found that simply see mindlines as heuristics, Gabbay and Le May explain that logical thinking is embedded in mindlines. What we derive from this is that the set of tools used to reduce or construct (depending on our perception of realities) the knowledge base for a clinical decision differs dramatically between conventional EBM and mindlines. In the former, the dominant tool for identifying knowledge is rational, conscious questioning and the main requirement of that knowledge is validity – that is, if it fulfils the criteria of correspondence (to the real world – for example through a robust sampling procedure), coherence (with what we define as a logical system of high-quality knowledge – for example, derived from a well-conducted randomised trial and meeting the standards of statistical prediction) and consensus (experts agree – for example through peer review). The literature on mindlines, as reflected for instance in the circular evidence model suggested by Walach et al. [63], fundamentally challenges these philosophical criteria

by which a finding will become classified as ‘true’ as they are inadequate to make predictions in the real world of clinical practice.

In particular, EBM intentionally focuses on so-called ‘frequentist’ reasoning, in its quest to discard mechanism-based reasoning and reliance on (potentially unreliable) clinical expertise [164,165]. I will discuss this type of reasoning again in chapter 6 in more detail, but in this discussion on mindlines it is important to know that this type of reasoning is an evasion of the fundamental inability to predict the future – or more precisely the problem of induction [28]. If we look at an association between A (such as a taking a tablet) and B (such as a clinical outcome) in the rich context of everyday practice, we may discover a web of interacting influences linking A with B. Evidence based reasoning is essentially a process of stripping the causal net (compare Pearl [166]) from the association between A and B (‘bias’), in order to find a single general yet virtual rule – and then applying this rule to predict the future in another situation where A and B, but also the bias, are present. This might be termed reality-to-rule-to-reality reasoning. In contrast, mindlines allow other evasions of the induction problem (such as Bayesian learning from a one-off experience) [28] in a chain of reasoning that might be termed reality-to-pattern-to-reality, which allows practitioners to keep the network of causality intact from one case to the next.

As articulated in the idea of system 1 and 2 knowledge [140], mindlines draw more on tacit knowledge, the knowledge we subconsciously use when focusing on the things we want to do [167]. Subconsciously knowing how to interpret gestures, smell, interaction, environment and time during a consultation reduces prevailing uncertainties and help

us to further shape our holistic understanding and make predictions of what is likely to happen in *this* case.

More importantly, mindlines encapsulate a more sophisticated and comprehensive concept of truth than traditional EBM. As Gabbay and Le May eloquently explain [15], mindlines take a constructivist approach to knowledge, assuming that it is created in social processes, through discourse, influenced by cultural and historic forces. This chimes with the work of the Russian philosopher Mikhail Bakhtin, who proposed that a unified truth involves sharing personal knowledge with others who provide a separate perspective. *'The idea lives not in one person's isolated individual consciousness—if it remains there only, it degenerates and dies. The idea begins to live, that is, to take shape, to develop, to find and renew its verbal expression, to give birth to new ideas, only when it enters into genuine dialogic relationships with other ideas, with the ideas of others'* [168]. Similarly, the Austrian philosopher Ludwig Wittgenstein argued by the same reasoning that all knowledge is collective [169].

Aligning with this focus on the intersubjective nature of knowledge creation [170,171], we contend that contrary to the fears expressed by Glasziou in the quote above, mindlines are not void of validation processes in spite of being mainly tacit. They convey strong and rich elements of shared sense-making (and hence consensus-making), both conscious and unconscious; they address correspondence with reality as it pushes back in the local context; and they address coherence using other types of evasions of the induction problem. In sum, mindlines can be accurate and useful in a local setting and

provide useful predictions, despite not being construed according the set of reduction tools and beliefs underpinning the EBM paradigm.

Economics, politics and ethics

The political, economic and ethical dimensions of the processes of knowledge creation in the papers in our sample were almost never directly linked to mindlines. This is surprising, given that mindlines and traditional EBM differ considerably in this perspective. Several papers in our sample noted that politicians, research leaders, management consultant firms, lobbying groups, the pharmaceutical industry and many other powerful actors use their influence to define research priorities, what counts as medical evidence, how knowledge is distributed and how resources are allocated [136,172–175]. In EBM, population-derived statistical estimates fit the needs of policymakers as they provide truths that are – apparently – 'right' for groups and those who interact with those groups, such as governments and the pharmaceutical industry. What is viewed as good care for a defined group as a whole is in some way regarded as good for individuals. Mindlines however lack this overarching 'built in' criterion of what is right or wrong patient care. With other authors, Gabbay and Le May worry that mindlines can spread 'collective folly' [15].

EBM strongly adheres to the 'deficit model' [176], which entails that clinicians and patients are regarded as deficient in certain knowledge: evidence-based knowledge. This is considered a moral problem of *'leaving people incapable of understanding the world around them'* [37]. Mindlines, on the other hand, correspond more with the idea that anyone, including patients, create valid knowledge too and can be 'experts' in

consultations [177]. With the current evolution towards person-based medicine and practices [178], the deficit model may be set to give way to a more pluralist and constructivist one. But at the same time this may uncomfortably question our basic assumptions about who decides what good or bad care is.

Knowledge management, knowledge intermediation

Although Gabbay and Le May aptly remark *'how ironically inconsistent it would be if [they would] try to dictate how [their] work should be put in practice'* [15], many of the implications they discuss have been explored in the literature we found. We discovered articles that explored which sources of knowledge clinicians actually use, projects that aimed to bring research and practice closer together and the development of organisational structures such as communities of practice and virtual social networks to support the use of 'knowledge-in-practice-in-context'. But what strikes us when considering these implications for practice is the question of whether controlling knowledge creation is actually feasible. EBM assumes that knowledge can be managed and that, through intermediation, the knowledge deficit of both practitioners and patients can be rectified.

In contrast, mindlines remind us by their emphasis on tacit knowledge that knowledge creation is in large part unmanageable. A one-off event experienced by an individual is all too real for them, 'anecdotal' though it may be to others. Furthermore, we cannot control all interactions, nor can we control all aspects of resources and contexts. Knowledge development is an organic rather than rational process, which can only be controlled to a limited extent. Currie et al note that the implication of mindlines is that

'any attempt at managing professionalised and tacit knowledge in health care through the mobilisation of explicit and codified knowledge faces significant challenge'[11]. As Contandriopoulos et al conclude in their review on the dissemination of knowledge: *'…the quest for context-independent evidence on the efficacy of knowledge exchange strategies is probably doomed'* [142]. This picture of knowledge as fluid, multiple, uncontainable and defying rationality is a long way from the hopes and dreams of the EBM movement [20] or even science itself. Thomas Kuhn touches the core of the matter writing: *'We are all deeply accustomed to seeing science as the one enterprise that draws constantly nearer to some goal set by nature in advance. But need there be any such goal?'* [14].

However, it would be wrong to conclude that because mindlines are constructed, tacit, emergent and shared, they are directionless. On the contrary, there is evidence that the knowledge of mindlines is 'self-organising', tending to achieve stability over time. Broekaert noticed *'real human commitment consists of an open, methodical, meaningful search for the best solution for a certain problem'* [179]. I will discuss theories of what is true and valid in more detail in chapter 7. Here, is important to realise that consciously and subconsciously, we collectively create and continuously refine more or less enduring frameworks to look at the world, based on our previous experiences, opinions of colleagues and experts, practical knowledge, guidelines and articles, produced in discourse, agreement and consensus with others, limited by psychological abilities, contexts and the physical world. The more closely the statement, 'This patient probably has a viral illness, tomorrow they will be OK, even if they do not take antibiotics today' persists in such a framework, the more real, useful and valid it becomes.

Persistence is not without risk. In evidence-based guideline development and research, we are used to synthesising a single version of reality to settle differences [18] and provide consistency of care. But 'consistency' of care can harm as well as help individual patients, since such an approach may limit our list of differential diagnoses to common or obvious options, removing the possibility of managing the unusual case differently [180]. Similarly, we have to prevent our mindlines from becoming too rigid. Unanimity, or the absence of logical contradiction, prevents the development of new, competing theories and innovations, which needed from time to time to replace the current paradigms [14].

This raises the question of why persistence of knowledge in the form of mindlines is valuable if knowledge is inherently ephemeral and too much persistence risks making our decisions too rigid. Ernst Mach argued: *'If our dreams were more regular, more connected, more stable, they would also have more practical importance for us'* [181]. Haridimos Tsoukas contends that *'[U]nderstanding presupposes an Archimedes' point, a perspective (undoubtedly an irremediably open-ended and evolving perspective, but a perspective nonetheless) from which the world may be viewed, accounted for, and interpreted. Ironically, abundantly available information leads to formlessness and, thus, to a diminished capacity for understanding.'* [182]. Persistence keeps a perspective open for exploration, prediction and guidance of human behaviour. We need persistence to help us to see and find new events, insights and practices so we limit discontinuity, or instability that nobody agrees with. Infinite reality is not created,

distributed, translated or mediated. We miss things because *we have to* in order to experience anything at all.

Hasok Chang [183] argues that scientific realism should commit to pursue many theories to find where reality 'resists', whilst investing to preserve theories that did not seem to work that well. In the future those might turn out to give helpful alternative insights. Applied to mindlines, this may translate to a call to create a broad menu of mindlines to find where collective reality 'resists' using many methods of truth finding. If we want to intermediate the process of knowledge creation (to the limited extent that this is possible), further research needs to look into how to speed up the cycle of building and turning over many more persistent mindlines, whilst keeping alternative, less persistent ones afloat efficiently.

4.4 Conclusion

In this chapter I argued that mindlines offer a philosophically and theoretically sophisticated perspective on knowledge and clinical method. Yet in ten years since the concept was introduced, the study of mindlines has remained a minority sport within critical social science while research within the EBM movement on the generation, circulation and use of evidence has appeared to remain predominantly (though by no means exclusively) wedded to a naïve rationalist view of knowledge.

That does not mean that the interest in mindlines is fading. A brief search on Pubmed in 2019 reveals that at least another 162 papers have been published that refer to

mindlines beyond 2015. Further research would be required to see if they reveal any new framings of mindlines.

In our sample Malterud noticed that the EBM movement does not limit the best evidence to RCTs and meta-analysis, *"Yet, the foundation for integrating the available sources of knowledge remains unclear. We still do not know whether convincing information leads to optimal decision making"* [184]. Similarly Jonas argues that *"we need to broaden and deepen our understanding of what counts as "evidence" and which types of evidence are best used to inform differing aspects of clinical decision making"'*[185].

In a similar vein, I further explore and develop epistemological assumptions underpinning EBM brought up in this chapter to see how knowledge is created and validated by both guideline makers and clinicians in the following chapters. The next chapter first explores the understanding of EBM as a rational and scientific authority and looks at the practice of 'purification' (relating knowledge with reality) through Latour's discussion of the concept of modernity.

Chapter 5: On purification: has evidence-based medicine ever been modern?

This chapter is an updated version of a paper co-authored with my supervisors called 'Has evidence-based medicine ever been modern? A Latour inspired understanding of a changing EBM', originally published in the Journal of Evaluation of Clinical Practice in 2017 [89].

5.1 Introduction

Many authors consider that the emergence of EBM (as a movement that continues to evolve) marks the shift from the traditional 'priestly authority' of doctors to a more modern, knowledge-based – and hence rational and scientific – authority [17–19]. As such, EBM/EBHC deserves analysis within the wider context of modernity in general. In this chapter, I consider the discourse on EBM and its future in the light of work by the French anthropologist and philosopher Bruno Latour, who analysed the emergence of science and technology in late modernity in one of his classic books 'We have never been modern' [21] . I will argue that evidence-based practice struggles with purification and EBM as a situated practice rather than as a sequence of research-driven abstract decisions [89].

5.2 'We have never been modern'

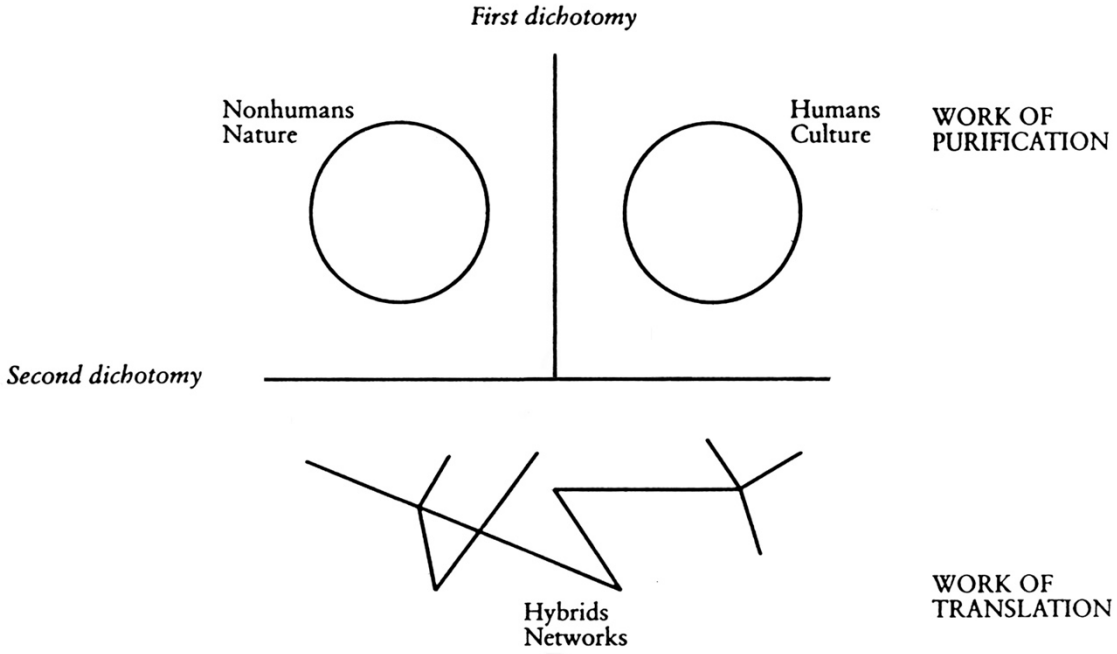
The term 'modernity' is generally used to describe a historic period in European arts, politics, science and the humanities that began in the 16th century with the rise of the new cosmology and experimental science and extended to the 20th century. It conveys the idea of a general progress towards the rejection of traditional and religious explanations as envisioned by enlightenment philosophers as well as early social theorists such as Max Weber, Emile Durkheim and Auguste Comte. Their work culminated in a broad range of related theories and ideas, some of which are still prevalent today. With modernity came increasing secularisation, rational thought and rejection of tradition along with the firm belief that reality is based on empirically testable scientific laws.

Bruno Latour wrote several classic books on the practices and nature of science that already inspired others to discuss relevant issues of the evidence based movement, such as the standardisation of morals and patients [186] and the implementation of innovations [187].

In 'We have never been modern', Latour acknowledges many versions of modernism, which is recognisable when society is engaged in two separate practices: *purification,* which means to make a distinction between nature (which has always existed) and human culture (which is man-made); and *translation/hybidrisation*, defined as efforts to combine elements of nature and culture in 'hybrids networks'. See figure 1.

Latour gives as an example of translation a debate on climate change, in which translation efforts seek to align elements of both nature and culture – specifically, "[the] chemistry of the upper atmosphere, scientific and industrial strategies, the preoccupations of heads of state, [and] the anxieties of ecologists".

FIGURE 1



Latour argues that modernists recognise that nature and culture have interconnections, but they seek to make them analytically separable. In particular, say modernists, science should be politics-free. Modernists' assumption (what Latour calls 'the modernist constitution') entails that nature is always objectively knowable and never man-made, and human culture is always made up and never directly influenced by nature.

According to this view, we would be 'modern' when and to the extent that we could successfully separate nature and culture and subject the former to rational scientific analysis. However, argues Latour, despite the application of advanced scientific techniques and technological apparatus, modern society is struggling and failing to keep these two phenomena apart. Nature and culture remain inextricably linked, just as they were in the pre-modern era.

To resolve this modernist problem, Latour does not argue for a return to pre-modern obscurantism, i.e. deliberately precluding full understanding of anything. Instead he introduces an alternative take on reality, which he calls a 'middle kingdom'. In this view, the world – consisting of both nature and culture – is explained with reference to 'hybrids' or so-called 'quasi-objects'. He offers a number of examples including soccer balls [188], frozen embryos, expert systems, digital machines, sensor equipped robots, hybrid corn, data banks and psychotropic drugs [21]. All of them are *naturally* real in the sense that they obey physical laws but are also *culturally* real in the sense that they have cultural meaning and social significance. They are objects and non-objective concepts at the same time. According to Latour, moderns deny quasi-objects even though they depend on them. To enter the middle kingdom means to acknowledge that nature and culture can never be kept apart and that the modern project is equally dependent on purification and translation.

5.3 Early EBM as a modernist movement

The emergence of EBM in the early 1990s can be understood as a modernist movement. For instance, an article by the EBM Working Group in JAMA in 1992 opens as follows:

> *"A new paradigm is emerging. EBM de-emphasises intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research."* [20]

In the above quote, real patients and conditions are understood as hybrids of two pure forms: objective nature and subjective culture. Clinical research is depicted as able to provide 'facts' of biology and epidemiology, as far as possible void of human intuition or experience. Indeed, the programme of training that was to be offered to the early students consisted (more or less) of a menu of tools, techniques and processes for separating objective (hence, trustworthy) evidence from subjective (untrustworthy) opinions, biases and perspectives. Thus, EBM began with explicit aspirations to modernism that still persist today [189]. For instance, in a paper on evaluating e-health interventions, the authors write:

> *"[..] health information systems should be evaluated with the same rigour as a new drug or treatment programme, otherwise decisions about future deployments of ICT in the health sector may be determined by social, economic, and/or political circumstances, rather than by robust scientific evidence."*[190]

Not all early protagonists of EBM rejected the value of clinical expertise. Indeed Sackett et al. famously defined EBM as an approach that integrates the best external evidence with individual clinical expertise and patients' choice [3]. But even this more accommodating definition remains essentially modernist. The fundamental issue here is the concept of "the best external evidence". Although Sackett et al advise practitioners to "integrate" clinical expertise and patient choice (which implies some kind of hybridisation) they persist in conceptualising evidence in the form of general truths that are external to the messiness of the clinical encounter and arrived at through rational, objective means (notably, randomised controlled trials).

Compare with Latour who writes:

> *"The moderns [..] did not make quasi-objects disappear by eradication and denial, as if they wanted to simply repress them. On the contrary, they recognized their existence but emptied it of any relevance [..] The modern explanations consisted in splitting the mixtures apart in order to extract from them what came from the subject (or the social) and what came from the object. Next they multiplied the intermediaries in order to reconstruct the unity they had broken and wanted none the less to retrieve through blends of pure forms."*

Thus, whilst Sackett and other early leaders of the EBM movement celebrated clinical expertise and the patent voice, they nevertheless insisted that evidence is developed in a secluded space outside the clinical encounter and *only then* translated through

integration with expertise and preferences. Thus they reproduced the dichotomy between science and translation typical of modernism.

By the late 1990s, critical social scientists were reflecting in an empirical study of attempts to achieve change in clinical behaviour across a United Kingdom Health Authority that EBM had succeeded in producing

> *"…[a] Cartesian epistemology of differentiation – a reality of distinct subjects and objects – in which research evidence and clinical practice are seen as dualistically opposed. Evidence, in this context, is seen as a commodity, whose better, quicker, easier to access, and increasingly electronic transfer, from the research pole to that of practice will lead to the more effective clinical management of patients […]. Whilst the appeal of this dualist model is its maintenance of stable boundaries and firm divisions: objective/subjective, research/practice, facts/values, such thinking is almost always structured so as to privilege one side of the dualism over the other. This has made possible the hierarchical distinction between, for example, the objective 'facts' of biomedical research and the subjective 'mere knowledge' of clinical practice. Here, an analytic and disembodied (scientific) organisation of knowledge is privileged over more tacit and situated experiences: a body of evidence, separated from its social context, that can be unilaterally transmitted from the research setting – where it is known – to the world of practice – where it is not."* [187]

The 1990s and 2000s saw an exponential expansion in the kind of research designated by the EBM community as "robust" – defined in terms of the methods used to reduce or eliminate bias in the collection and analysis of empirical data. The large randomised controlled trial became the bread-and-butter type of research and in its trail rapidly followed increasingly sophisticated systematic reviews and meta-analysis, instruments to appraise clinical practice guideline development, frameworks to guide and assess clinical care pathways, and the institutionalisation of implementation as a science.

5.4 The 'movement in crisis': EBM's receding modernist vision

The original vision of the EBM movement was that within a generation, the evidence base for clinical decisions would have been collected, collated and distributed for easy access by clinicians and policymakers. This knowledge base – "best evidence" – would be objectively verifiable and readily updated in real time as new research evidence from robust study designs accumulated.

However, research evidence in many fields remains heavily contested. The scientists, clinicians and guideline developers involved are usually EBM experts, they understand the rules for interpreting clinical trials and screening programmes perfectly well. Yet they each interpret the evidence differently. These disagreements between experts often originate not from the data but from a difference in *values* [65]. Principles like confidentiality, equality and preventing harm have an impact on how 'facts' are understood. Re-interpretation of existing evidence in national breast cancer screening

programmes based on mammography caused heavy debate about their effectiveness and the risks of over-diagnosis [191]. The discussion on the harms and benefits of statins in low risk patients and the elderly is fierce and ongoing as some trialists and systematic reviewers are accused of conflict of interests [192]. And the medical community was shocked when recommendations for Tamiflu [193] and paroxetine in adolescence [194] were reverted after previously unpublished research data was disclosed and re-interpreted by others. Again and again evidence-based value-free 'facts' turn out to be value-laden (and for good reason).

In the light of these examples, the warning offered in 1997 by Haridimos Tsoukas was remarkably prescient (though few in the EBM movement heeded it at the time):

> *"Contrary to how knowledge was viewed in pre-modern societies, knowledge now tends to be understood as information, that is as consisting of objectified, commodified, abstract, decontextualized representations. The overabundance of information in late modernity makes the information society full of temptations. It tempts us into thinking that knowledge-as-information is objective and exists independently of human beings; that everything can be reduced into information; and that generating ever more amounts of information will increase the transparency of society and, thus, lead to the rational management of social problems. However, … the information society is riddled with paradoxes that prevent it from satisfying the temptations it creates. More information may lead to less understanding; more information may undermine trust; and more information may make society less rationally governable."*[182]

The original goals of EBM to provide a methodology to generate independent truths and objectively assess the certainty of facts have so far proved elusive. Society's influence pervades. Despite the hopes and predictions of its forefathers, EBM has never become modern.

5.5 EBHC – moving towards a 'middle kingdom'?

A number of recent developments suggest that the EBM/EBHC movement is gradually recognising its rigid adherence to modernism while it attempts to get to grips with its stance towards nature on one side and human society on the other.

1. The discourse on EBHC displays a greater degree of reflexivity from within the community. Whereas the early protagonists of EBM saw the movement as part of the inevitable march of scientific progress, social scientists have always viewed EBM as historically – and politically – situated. Judith Greene [195] for instance notes that:

> *"The EBHC movement has attracted critical comments from sociologists and policy analysts from a number of perspectives: that for instance, clinical epidemiology is not always the most appropriate framework for decision making in healthcare [..] and that EBM can be deconstructed as a discourse which functions to maintain the 'purity' of medical practice in the face of threats to autonomy [..]. [O]ne key problem within the EBM*

movement lies in its assumption of rationalistic behaviour and linear change: that evidence is stable and independent of social relationships."

But now the EBHC community too appears to embrace these criticisms as it tries to find solutions for the inability to fully be modern. For example, in order to deal with vested interests the ALLTrials campaign [22,23] calls for better registration of planned trials, summary results reported within a year after the end of all trials and publication of a full report including all harms. These may still be considered modernist, purifying activities. But the practices of the initiative, creating momentum, networking, lobbying, beating the pharmaceutical industry at its game, are not. They make evidence-based minded clinicians, researchers and guideline developers more and more aware that medical 'facts' are not so objective as they seem. They turn out to be constructed, situated within specific agendas and alignments of key players in human society.

2. The EBM/EBHC community no longer views the result of a sample mean in a RCT as applicable unproblematically to the single case scenario. There is a more sophisticated understanding of the translational process from scientific evidence to a particular situation such as a complex intervention at the organisational and system level, or an individual patient in a particular situation.

For example, inspired by the concepts of 'Shared decision making' [24] and 'person centred care'[25], guideline makers are supporting individualised clinical decisions by providing new tools in new formats, as for instance the Magic app [196] or Option grids [26].

What is new about these instruments is that they do not just prescribe what objective, purified nature says (does the drug work?), but also present arguments from human culture (Do patients value it? Is it affordable?). True, as they still separate nature and human society in neat categories, they remain largely modernist. But nature and society are put on a more equal footing than ever before.

These developments are still somewhat removed from what it takes to care for an actual patient. That would be what philosopher Annemarie Mol calls a 'logic of care' – a constant tinkering, while avoiding neglect [31] or a *relationship*-based care – following the story of the person in context, in the words of Schei [197]:

> "*a relational competence, where empathic perceptiveness and creativity render doctors capable of using their personal qualities, together with the scientific and technologic tools of medicine, to provide individualized help attuned to the particular circumstances of the patient."*

But they do reveal a growing interest in the EBHC discourse to start thinking from the standpoint of real patients in real encounters as with Latour's quasi objects.

3. There is an increasing interest in less purified kinds of knowledge, research methods and styles of reasoning [28].

For example, work by Gigerenzer [70] on heuristics and other researchers on gut feelings [198] provide evidence that decisions based on not-so-explicit human reasoning and inferences does give good, and sometimes even better, outcomes in reality. Ethnographic research by Gabbay and Le May and observational research by Zwolsman et al. [10,149] showed that in everyday practice most clinical decisions commonly rely on tacit, hard-to-explain knowledge, influenced by past experiences, peers and contexts. Pragmatic randomised trials have been designed to test interventions in the *actual* context of where they will be used [199]. And guidelines makers urge that in practice good recommendation require more kinds of knowledge than just RCTs [2].

These examples show an understanding that the products of modernist practice of purification (resulting in explicit law like rules of medicine) need to be watered down to become applicable 'in reality'. Latour writes:

> *"We do not need to attach our explanations to the two pure forms known as the Object or Subject/Society, because these are, on the contrary, partial and purified results of the central practice that is our sole concern.*

> *The explanations we seek will indeed obtain Nature and Society, but only as a final outcome, not as a beginning."*[21]

> Purified findings from well-designed RCTs and reviews are wonderful insights but at the end of the day they may prove useless in the face of an actual patient, in the here and now.

Of course, consultations themselves have always been about translation and hybridisation. In consultations patients and healthcare professionals swiftly connect molecules, plaques, platelets, cells, average blood pressures, drugs, lifestyle, food, function, meaning, values, sense and care. What is new is the growing acknowledgement in the discourse on EBHC that these are not clearly separable in objective nature or subjective culture (as critical social scientists have argued for decades).

5.6 Discussion

In his book 'We have never been modern', Latour sought to eradicate the divide between science and culture. His middle kingdom implies a fundamental questioning of the modernist conception of science that extends beyond EBM.

Whilst current initiatives in the EBHC movement may concede the existence of a middle kingdom – that is, they accept that research evidence cannot sit above history, culture or politics but is recursively linked to all three [200]– there remains an important

paradox. As Latour argued about science in general in 'We have never been modern', EBHC cannot acknowledge this middle kingdom without ceasing to be modern and collapsing back into pre-modernism.

This creates an ambivalence: on the one hand, EBHC (at least as it has developed) depends on mediation, translation or a 'cultural supplement'. On the other hand, it still relies on the purification of science to be separate from, and more objective than, real-world messiness. The whole concept of "best evidence" hinges upon such a divide. If not, it risks becoming 'unscientific'. Inevitably, then, EBHC must move towards a middle kingdom to deal with the paradoxes it has already generated (such as the dominance of vested interests in the publication of large randomised controlled trials in mainstream academic journals). But EBHC can never fully embrace this middle kingdom without effectively rejecting the very concept of evidence on which its central claims are based. Instead of resolving this paradox, the EBHC movement must co-exist with it and 'muddle through'.

Latour argues that whilst modernity is built on a myth (the separability of nature and culture), it nevertheless produces benefits. Modernity generates the need for translation, hybridisation and cultural/technological supplements, a need that follows from the act of purification itself. The practice of purification also leads very efficiently to many new insights (smoking leads to cancer), concepts (statistical significance) and natural objects (DNA). All these 'quasi-objects' would not have existed in a pre-modern society, he argues. Indeed, it is hard to underestimate the impact EBHC has had with its

renewed methods of induction as means to purify clinical reality into more or less stable rules and recommendations.

The AID GIN working group contends that any new interpretation of EBHC should not seek to refute the fundamental principles of clinical epidemiology or the position of the RCT as the best study design for reducing bias in human experiments. Guideline makers from the group are not arguing for a return to the priestly authority of doctors, to anecdotal evidence or to guideline development by GOBSAT (good old boys sat around a table). Rather, they suggest that a progressive, mature, 'middle kingdom' EBHC would reframe the concepts of 'evidence' and 'medicine/healthcare' in EBM / EBHC in several different ways:

- The new EBHC should reflexively reject the science-culture dichotomy. Especially in some circumstances (for example, where different stakeholder groups hold widely different values and priorities; where vested interests loom large; or where uncertainty is inherent [104]), 'pure' evidence is unlikely to be obtained by stripping away real world messiness or imposing rigid or unworkable protocols. As Contandriopoulos et al point out:

> *"Collective knowledge exchange and use are phenomena so deeply embedded in organizational, policy, and institutional contexts that externally valid evidence pertaining to the efficacy of specific knowledge exchange strategies is unlikely to be forthcoming* [172]*".*

Rather, a mature EBHC will seek to incorporate the messiness of real world hybridisation into a wider range of approaches and methodologies. Concepts of a 'gold standard' and a 'hierarchy of evidence' become highly problematic in such circumstances. RCTs require 'supplements' to create better inferences for a complex reality, be it pragmatic trials, pattern recognition modelling, participatory co-design or other approaches.

- The new EBHC recognises that all evidence – whether derived from science on the one hand or from human culture (patient preferences, opinions, insights, etc.) on the other – requires translation and interpretation. This is in contrast to a recent study where researchers found that guideline panel members differentiated between 'objective evidence' from studies and 'insights' from patients. They write: "There is a tendency to try to weigh-up the patients' views *against* the clinical and economic data, which proves difficult." [201]. With Latour's framework it becomes clear why: nature and society are hard to separate. If we accept the legitimacy of the middle kingdom, we must accept that research evidence cannot be placed on a pedestal apart from other evidence in this way, as well as accepting that research evidence will always be developed *outside* the relationship-based care for the patient and based on a logic which is different from the relational logic inherent in a clinical encounter.

> *"The danger of the dominant discourse of rationality is that by marginalising and devaluing the role of individual judgement, it undermines rather than strengthens actors' capacity to act"* [202]

- The new EBHC acknowledges the need to reason not only from the individual to the general or from the general to the individual but from individual to individual (casuistry [203]). In the old EBM healthcare needs to be purified to find its (assumed stable) laws of nature. For instance, people with hypertension have a higher cardiovascular risk. We, the society of humans, make mistakes against the absolute rules of nature which are typically referred to as 'cognitive biases'. Instead, Latour urges us to reason from the original ‘quasi object / hybrid’, that is the patient in practice in clinical context. The question then becomes: is this patient with hypertension one of the group of patients that would *form* the 'rule’ (hypertension causes increased cardiovascular morbidity)? It is a small difference, but crucial in EBHC: it forces clinicians to challenge the rule with every patient they see. To ask if there actually is a rule or an original new situation. And to recognize that the rule found in a certain reference population (e.g. incorporated into a guideline) cannot remain a law forever. It will change as the original/real patients in their contexts are changing. It is about causality departing from a dynamic world context, not a stable controlled laboratory environment [204].

- The new EBHC recognises its own limitations in relation to decision support and inductive inference in healthcare.

Latour's framework unveils EBHC as just one of the activities (practices) of everyday healthcare. It is the practice of purification in that realm: making inferences to produce new insights and good decisions.

But as a result of the focus on purification practices of early EBM, the purpose of healthcare itself became skewed towards aiming for objective evidence from nature at the expense of care and equity. The concept of healthcare as caring, loving and nurturing has largely been lost in the process [4].

This was not intended. For Archie Cochrane, who the early EBM proponents often cited in support of their view, health care had two equality important dimensions:

> *"I see the NHS, rather crudely, as supplying on the one hand therapy, and on the other board and lodging and tender, loving, care"* [205].

Purification and hybridisation have their place, but not predominantly. Healthcare is about making good inferences (which is the role of EBHC), *and* about equity *and* about relationship-based, loving care. Paraphrasing Latour: healthcare should once again be reinstated as broad concept that has both "science and society as its satellites".

5.7 Conclusion

In this chapter I considered the current position, changes and the future of the evidence-based movement by introducing Latour's views on modernity as a framework for discussion. It introduced the role, the paradox, benefits and harms of the modernist practice of purification, which links knowledge and reality. In the next chapter I explore epistemological challenges that guideline makers face in their work and offer a perspective on clinical reasoning in evidence-based decision making to make inferences about future reality.

Chapter 6: On inference: evading the problem of induction

This chapter is an updated and adapted version of 'Different knowledge; different styles of reasoning. A challenge for guideline development', a paper originally published in BMJ Evidence-based Medicine in 2018 [90]. *This paper was developed in collaboration with the GIN AID working group.*

6.1 Introduction

Although evidence on statistical association – usually from RCTs - is commonly thought to be the dominant type of knowledge appraised and included, guideline developers frequently use a range of other types of knowledge including the views and experiences of those using and providing health services, understanding of how interventions work (for example, from logic models or realist evaluations), and other information, such as aetiology and the context of care [2]. This would include, but not be limited to, a broad range of explicit understandings of knowledge such as knowledge from outbreak investigations, laboratory research, mathematical modelling, qualitative research, quality improvement processes and clinical audits but also tacit knowledge, practical knowledge and heuristics.

These different types of knowledge are used and needed in many situations; for example, when evidence from RCTs is not available, impossible to obtain, contradictory, or inappropriate. They can also be used in conjunction with knowledge from RCTs to

provide context, to assess relevance and to understand bias. Furthermore, not just explicit (written or spoken) knowledge, but also the more intricate forms of knowledge like experiential and contextual knowledge can help guideline makers to take an approach consistent with the intentions of early evidence based medicine (EBM) proponents: namely, that best evidence is not restricted to evidence from RCTs and meta-analyses alone [3].

However, how to properly appraise (judge) and include (integrate) different kinds of knowledge remains unclear. Agreed methods are not yet available or are in the early stages of development and the need for and use of different kinds of knowledge is not always explicitly acknowledged which affects the use of guidelines in practice [206,207]. International and cultural differences in guideline production practices may further impede developments in appraising and including a broader range of types of knowledge.

Guideline development methods have been developed by international networks and presented as universally applicable. But there may be differences in what methods are most appropriate for different contexts. In the workshops and conference sessions organised by the AID Knowledge Working Group, guideline developers from Europe, Latin America and Australia have generally welcomed efforts to include a wider range of types of knowledge in guidelines. USA-based guideline developers have been more critical, especially voicing the idea that any guidance not based on the highest level of evidence derived from RCTs should not be called a guideline. Such different assessments

of what knowledge can / should be included in guidelines may have to do with the context in which guidelines are developed, including factors such as how the health care system is organised, who develops a guideline, litigation, what additional risks for conflict of interest this implies, and therefore whether guideline methods are more rigid. The working group finds that approaches towards guideline appraisal and inclusion need to suit the context in which they are used.

In this chapter, three specific aspects of guideline development are discussed to highlight the main challenges identified in collaboration with the AID Knowledge Working Group through discussions and workshops with guideline developers and users:

1. the purpose of guideline development,
2. the problem of induction,
3. the dominance of frequency-based reasoning, and
4. the challenge of integrating different sources of knowledge.

6.2 The purpose of guideline development

The efforts of the pioneers of the EBM movement were primarily in response to the discovery of the variation problem in population studies. Reducing variation of the care provided at a population level was considered to be an important way to achieve improved quality for individual patients [208]. Hence epidemiology, the science of studying *populations*, gained prominence in guidelines, the aims of which are to support

decisions for *individual* patients. Classic epidemiology became clinical epidemiology when introduced to the bedside, and the dominance of RCTs as the gold standard for intervention studies to assess causal relation between interventions and effect followed in this construct of epidemiology as used in EBM. The underlying - yet little explored - assumption is that guidelines based on population studies provide the best advice to inform clinical decisions for individual patients or situations.

However, reducing variation is not the only reason for developing guidelines; they are developed for several reasons, of which the most important one is to improve the quality of care. In order to meet the range of needs, guidelines may need different approaches; such as summarising large quantities of knowledge for practising healthcare professionals, serving as an intermediate product for other tools or applications (such as clinical decision support software) or providing implementation guidance. Although not primarily developed for this purpose, guidelines can also serve as tools to legally shield both patients and professionals, to help governments and health insurers allocate scarce resources and to act as governance frameworks for practitioners and governments.

There is also the role of guideline development as a discipline in itself; along with its associated practices and institutions, it provides employment and intellectual interest for many. There has been surprisingly little research into the purposes of guideline development. One mixed-method study found the purposes of guidelines were: defining

norms, summarising evidence, formalising current consensus and/or describing current practices in a handbook-type format [206].

Making the purposes of guidelines more explicit may help determine how different types of knowledge could and should be used. For instance, if the aim is to describe current good practice, (for example, how services are organised to deliver care) this may be better achieved by drawing upon qualitative or mixed-method evaluative research rather than RCTs. If the aim is to assess the effectiveness of a specific treatment or approach, evidence from RCTs or high-quality prospective cohort studies would usually be the primary source of knowledge, with qualitative or mixed-method studies serving to help understand the local context of implementation.

It is important to note, however, that for most guideline developers, the primary purpose remains that of supporting decision making in the clinical encounter. This leads us to the next fundamental aspect of guideline development.

6.3 The problem of induction

How do different types of knowledge in guidelines development help to make clinical decisions? Some basic concepts from the philosophy of science may help to understand the problem.

In logic, to *infer* means to conclude from evidence using reasoning [28]. In everyday healthcare practice, care professionals and patients reason to reach conclusions about what has happened, to make predictions about what will happen and to decide what to do next. Because of uncertainty in medicine, we usually deal with a specific type of inference, called induction, where the conclusions of our reasoning are not always right even when based on true premises. In philosophy there is a concern whether this is actually possible, called the problem of induction [28], as introduced by Hume in 1739 [27]. At its simplest this means we cannot predict the future with certainty. Although this seems reasonable, we are in fact able to predict the future quite accurately on many occasions in clinical practice. How is this possible? Philosopher of science Ian Hacking [28] argues that we never solve the problem of induction, but only evade it by applying different kinds of reasoning to reduce uncertainty and increase our chance of reaching the best possible outcome. zz

6.3 The dominance of frequency-based reasoning

The evasion most dominantly used in guideline development is frequency-type reasoning in the form of systematic reviews, RCTs and observational studies [206]. This evades the problem of induction by recognizing that “although we cannot predict the future for the individual case, we can be 'usually' right (e.g. 95% of the time)” [28] as long as events or cases are frequent enough. The central reasoning in RCTs can be summarised simply as assuming that everything in the world (such as people, relations) behave like dice. That is, if dice roll perfectly, they are fair. If they are unfair – where one side is heavier than the others – they are biased. There are of many types of bias a priori

in real life situations: for example, known causes, confounders and differences between groups.

An RCT of patients with therapy A versus patients with therapy B is like trying to make two similar dice as fair as possible, rolling them many times, and comparing the results of the rolls to show their tendency. The assumption is that by rolling two fair dice, one would expect to see similar results in the end. If this does not happen then the reverse must be true; one of the dice is unfair or biased; one of the faces is heavier than the other. In that case one of the therapies was better at causing the outcome. We do not know 'why' (rolling dice does not give us an explanatory mechanism), but we conclude that 'it just did', because we found an a priori 'unknown' bias.

It is important to note that frequency-type reasoning does not work for single case scenarios. A patient can get a first hip replacement only once, not a hundred times. There is no frequency about it. It's like throwing a dice once: one can never be sure what will happen. You either throw a six or you do not.

Frequency based reasoning relies on basic assumptions, that have some drawbacks. First, this line of reasoning assumes that reality is dice-like and that we – e.g. scientists, guideline developers and healthcare professionals – are rolling the same dice. Frequency-type reasoning pre-supposes adequate framing and defining of what is similar and what is not, which is always based on judgement and choice.

Secondly, frequency-based reasoning aims to find simple causal correlations, independent of context. The question is whether these simple correlations hold true in real life. Different understandings of causality exist that could help us address this drawback [204]. For instance, a network of complex causal relationships may be more realistic. This drawback is described as the *efficacy paradox*, where the different interference from non-specific effects (different from those controlled for between groups in a trial), measurement artefacts (that mimicked therapeutic effects in the trial) and regression patterns (such as the self-limiting nature of a disease) in real life can outweigh the specific effect found in a trial. This paradox may become especially apparent when inferring in the context of multi-morbidity [209].

Finally, and most importantly, although frequency-based reasoning works well for frequent events (large groups, many data points, long periods of time), such reasoning faces fundamental limitations when inferring in the single case scenario: a single patient, a rare disease, a system intervention, a one-off event. This can be particularly challenging when recommendations based on frequency-type evidence alone are deployed to help decision-making for individual patients or unique situations, such as a public health response to a disease outbreak [2].

Given these drawbacks, it is worth noting that other types of reasoning to evade the epistemological problem of induction exist. In table 6.1 at the end of this chapter several alternative ways of reasoning, or simply “evasions” according to Hacking, are listed. The list was conceived by the authors of the GIN AID Working group through discussions, workshops and considering the literature. The evasions are mainly used in areas where

frequency based reasoning is particularly problematic, for instance in guidelines focussing on complex interventions, public and occupational health, rehabilitation and social care and welfare [9,209,210].

These different types of reasoning try to help make valid inferences for the single case scenario when there is no frequency of events. Many of these are already recognised and stated by Bradford-Hill [211] in his criteria for causation, but some are newer; such as Annemarie Mol's concept of 'tinkering' [31], where a practitioner will try something, wait and see and let unfolding events guide the next step. Using this type of reasoning, the problem of induction is solved through 'tinkering'; making incremental changes to improve a situation.

Guidelines can and do support these kinds of evasions by including different types of knowledge. For instance, providing laboratory information about aetiology helps to make an inference based on mechanistic reasoning [212]. A description of cases of harm can offer an inference based on the precautionary principle [213]. Rethinking how inferences are made in practice, may shift the dominance of frequency-based reasoning and its reliance on a restrictive type of knowledge to a broader spectrum of knowledge being used to support different reasoning approaches. The need for using different type of knowledge is shown by a large Dutch analysis showing that knowledge from RCTs far outweighed other knowledge types used, irrespective of the question at hand, thus ignoring important and relevant knowledge from other sources [206,207].

6.4 The challenge of integration

Making a recommendation for a specific healthcare problem in a specific healthcare system requires the assessment of knowledge not just on its own merits, but importantly its integration with other knowledge. Indeed, EBM is defined as integrating the best evidence with clinical expertise and patient preference [3]. But in the context of medicine, and even more so in that of guideline production, integration of different types of knowledge remains under-explored and under-theorised. Some areas of evidence synthesis have addressed integration. For example, statistical techniques such as meta-analysis can be used to combine data from different studies; and another range of techniques can be used to synthesise qualitative data.

In guideline development, most of the activities and tools to support high quality evidence synthesis such as risk of bias assessment and quality assessment (such as GRADE) tend to focus primarily on frequency-based reasoning and knowledge. For the assessment of quality of qualitative evidence there are limited, but relevant initiatives for guideline development in progress; for example, the recently published GRADE-CERQual (Confidence in the Evidence from Reviews of Qualitative Research) [214] guidance. But many of these efforts try to achieve integration by appraising all knowledge again in frequentist terms, like with some qualitative evidence synthesis methods [215], that "emphasize frequencies of the qualitative data they present … undermin[ing] the uniqueness of the qualitative knowledge they proclaim by focusing on frequency and the general patterns" [216].

The main issue is that these tools, activities and initiatives aim to integrate *similar* knowledge, such as data from the same study designs, the same populations, or the same outcomes. How different kinds of knowledge are valued, appraised, and weighed in relation to each other, for example regarding effectiveness, efficiency, or ethical concerns, is not clearly articulated.

Nonetheless, guideline developers do recognise that other types of knowledge are often used and somehow integrated in practice, particularly when discussing the evidence and formulating recommendations, often called 'judgement' or 'considered judgement' [217,218]. This is the traditionally less clearly described or analysed black box part of the process which new initiatives try to shed a light on, such as NICE's structured tables linking evidence to recommendations and the GRADE Evidence to Decision frameworks [217].

They appear promising, yet challenges remain. Firstly, findings from ethnography question whether structured frameworks really influence or reflect guideline development processes [106,219]. In an ethnographic study of guideline development meetings discussed in Chapter 3.5, Moreira showed that guideline developers formulate guidance by combining different "*repertoires of evaluation*, organised around four different epistemic criteria: robustness, usability, acceptability and adequacy" [106]. Importantly, such criteria are deployed at each stage of evidence appraisal: usability, acceptability and adequacy are integral to evidence assessment, rather than being easily categorised as either 'judgements' or 'additional considerations' as current evidence to

decision frameworks suggest [217]. Acknowledging the importance of these epistemic skills in evidence appraisal become much more important when it is understood that recommendations nearly always draw upon different types of knowledge.

Secondly, bringing knowledge together is not just a process of integrating, triangulation and finding a single answer. Knowledge from many sources is often conflicting, and indeed the exploration of opposing ideas is often very important. In social sciences, methods for evidence synthesis of other kinds of data have been developed and assessed in research [158] and in practice guideline development [9], but these have not yet been adopted routinely in healthcare guideline development. A process of integration is not just a technical, simple mechanistic process. Guideline development is a human, social process involving relevant stakeholders in discussion, debate and judgement. Therefore, the guideline development processes also relies on a balanced and representative guideline committee that functions well [220].

Finally, guideline makers note that integrating many types of knowledge is not a process in which anything goes. Some integration processes are likely to be better than others. Guideline development needs to be transparent and consistent so that reality, be it physical or social, can limit the inferences and recommendations made. A range of integration approaches depending on our understanding of what is true and real is needed; for example, integration of different knowledge could be based on combinations of coherence (what fits best in a network of other theories), on consensus

(what people agree on) and/or on correspondence (what links best to what is believed to be real).

Given the current state of evidence to decision frameworks, there is still little guidance on how to robustly and consistently combine knowledge of different types without using the frequentist understanding of knowledge. Yet, guideline makers of the GIN AID working group find that the development of guideline recommendations is an interactive human process that requires a range of knowledge and experience including, but not exclusively, knowledge from frequency-based research, such as clinical trials. As in the clinical encounter, appraising and including different types of knowledge in guideline development should be used to make better inferences to guide decisions, but in practice arguments are used to exclude some kinds of knowledge for a range of reasons, including concerns about introducing bias in frequentist reasoning.

6.5 Conclusion

In this chapter I have presented and discussed epistemological issues in collaboration with the GIN AID Working group. It highlights important issues in guideline development regarding its purpose, use of different knowledges, reasoning and integration. It also reveals views on what guideline developers consider possible directions on how to cope with these challenges, including being more aware of context and incorporate a more pluralist (and less rigid) epistemology of EBM. I will address these challenges again in my empirical work in chapters 8 and 9, after delving deeper into bias as a concept that is

often used as an argument against using different kind of knowledge and reasoning in the next chapter.

Table 1 Types of reasoning to evade the problem of induction

Please note this table aims to provide only a brief overview of alternative types of reasoning and by no means covers the extensive literature on this topic. Types are not exclusive and overlap.

Type of reasoning (with examples of key scholars)	**Shorthand description**	**Brief explanation**
Frequency type (Neyman, Hacking)	A series of future events will be like a series of past events.	This type of inductive inference reflects Hume's proposition that we cannot be correct in our predictions every time, but it argues that with large, high-quality studies we will be right most of the time (say 95% or 99% or 99.9%) about series of future events [28]. It is impossible to talk about the probability of a single event in a single patient: there is no frequency.
Bayesian (Bayes, Hacking)	Updating degrees of belief	This type of reasoning also recognises that we cannot predict future events exactly but holds that we can learn from our experiences reasonably well. With the input of other types of reasoning, such as frequency type reasoning and mechanistic reasoning, we can update a prior probability (a degree of belief) not only about series of events but also about an individual event in a single patient [28].
Abduction (Peirce)	Reasoning to the best explanation	Abduction is making a prediction while answering the question "what, if it existed, would best explain what I have directly observed?" Whereas the Bayesian evasion takes prior probabilities as a given and is more about confirming them (or not), abduction allows us to imagine wider contextual influences and causal mechanisms about our prior beliefs even when we do not directly measure them [29,221].
Mechanistic	How things work	This type of reasoning makes an inference based on a mechanism such as a pathophysiological rationale or root cause analysis. Illari et al. define a mechanism as consisting of "entities and activities organised in such a way that they are responsible for the phenomenon" [212].
Falsification (Popper)	Trial and error	Popper [30] agreed with Hume: we cannot say anything about the future, there are only theories that cannot even be proven. At best, we can only prove that they are wrong (falsifiable). This 'anti-inductivist' reasoning suggests continuing using a certain theory, hypothesis, diagnosis, practice etc, until it is demonstrated that it fails.
Precautionary principle	In case of uncertainty about the future prevent harm	This type of reasoning, often used in environmental decision making and occupational health, favours to take preventive action in the face of uncertainty when making an inference. It puts 'the burden of proof to the proponents of an activity; exploring a wide range of alternatives to possibly harmful actions; and increasing public participation in decision making' [213,222].
Means-to-ends reasoning	Find many ways to reach a goal	This type of reasoning asks the question what ways there are to reach a wanted outcome and which of those ways would be the more efficient? It is about taking multiple measures to reduce the uncertainty that a certain outcome will be achieved (making our inferences more likely to become true) [223].
Tinkering (Mol)	Taking care while an uncertain future unfolds	In this view healthcare is more like a 'practice' than it is about making rational choices [31]. This approach puts emphasis on the importance of taking good care for the patient and the prevention of neglect. Inferring is a process that unfolds over time, while addressing many factors on the way. It is taking very small steps, limiting the amount of time between a prediction and the outcome.
Non-analytical reasoning (Gigerenzer, Stolper)	Using intuition	Non-analytical reasoning such as heuristics and gut feelings (combination of heuristics and emotions [198]) used to make inferences. These types of reasoning are considered fast, intuitive and automatic thought processes. Non-analytical reasoning can in certain environments outperform analytical reasoning in psychological, biological, sociological and economic inference tasks [70].

Chapter 7: Rethinking bias and truth

This chapter is an updated and adapted version of 'Rethinking Bias and Truth in Evidence-based Healthcare', a paper co-authored with my supervisors originally published in the Journal of Evaluation of Clinical Practice in 2018 [91].

7.1 Introduction

In chapters 4, 5 and 6 I discussed mindlines as an alternative way to understand knowledge, the challenging practice of purification and how guideline makers struggle with the dominant style of inference in classic EBM. In this chapter I discuss issues regarding assumptions of the EBM paradigm on how to establish whether knowledge is valid and true.

Truth is and has been one of the most central and extensively debated topics in philosophy [224]. Whilst philosophers are keen to understand the nature of truths and what makes them true, scholars of evidence-based healthcare (EHBC) however have tended to more interested in bias. A simple word search on PubMed gives about 181,000 hits for bias and just 22,000 hits for truth. Epidemiologists generally address the topic of bias by describing categories of kinds of bias and offering suggestions on how to minimize or eliminate each [225–232].

The focus on research methodology (specifically, the practicalities of experimentation) and clinical application (specifically, poor reasoning in the clinical context) may go some way to explaining why bias became so important in evidence-based practice and hence why truth as a concept was largely neglected. As a result, however, the dominant discourse of bias in EBM is inclined to wrongly normalize truth as a given. It overlooks that the debate around truth is far from resolved and indeed remains a live topic in philosophy and the humanities.

Also, the way bias is understood in EBM does not seem to provide sufficient argumentation against alternative facts, factoids and post truths. Renewed concern of falsities being propagated on an unprecedented scale through social media has inspired many from outside [233–235] and within the research community [236,237] to come up with new solutions to counter those. But these arguments tend not to address the more fundamental question: how do truths differ from untruths (and what is the nature of the grey zone in the middle)?

Most importantly, the discourse on bias in EBM assumes that truth is unproblematic and that, therefore, the right decisions will emerge once all sources of biases are defeated. This flawed assumption constrains thinking of what good decision-making in the clinical encounter actually entails.

Belgian philosopher Isabelle Stengers offers an alternative view in her books ‘Sciences et Pouvoirs’ [238] and ‘Power and Invention’ [32]. She sees bias as productive and necessary to forge facts. Drawing on her work and that of others, we set out in this paper

to uncover some of the fundamental beliefs and assumptions on bias and truth that drive and constrain what we can know and do in clinical practice. We conclude that much could be achieved by turning our attention to discussing, teaching and extending theories of truth and considering its relationship to bias in evidence-based practice.

7.2 The ideal limit theorem of truth

The philosophical debate on truth spans several millennia. However, contemporary authors on the subject usually address a set of dominant views in one way or the other [239–242]. Correspondence theories of truth hold that what is true should somehow represent how reality actually is. Coherence theories see truth as what coheres with a whole set of beliefs. Pragmatist theories of truth refer to what works in practice. Constructivist theories of truth are concerned with how scientists interpret the world and how particular interpretations come to shape research traditions and empirical choices [243]. Deflationist theories attribute limited significance to the concept of truth and question what it actually means to say something is true. These broad philosophical positions are by no means exhaustive; many other views on truth exist in the literature.

We suggest that the epidemiological perspective taken by most EBM scholars implicitly is a concept of truth known as the ideal limit theorem [28,239] as commonly attributed to late 19th century American philosopher C.S. Peirce [29]. Although Peirce propagated several concepts of truth [244], he is most famous for the belief that the truth would reveal itself in the long run, over many events in many communities now and in the future. He states:

> *"Truth is that concordance of an abstract statement with the ideal limit towards which endless investigation would tend to bring scientific belief, which concordance the abstract statement may possess by virtue of the confession of its inaccuracy and one-sidedness, and this confession is an essential ingredient of truth* [245]*"*

This idea of the long run finds its way to healthcare research via Fisher's publications in the 1920s on scientific experiments [246] and Bradford Hill's work on clinical trials in the 1930s [247] which led to the important methodological development of the randomized controlled trial (RCT) in which participants were allocated to treatment and non-treatment groups with the explicit goal of eliminating bias [248]. In the Lancet in 1937 Bradford Hill writes:

> *"The reason why in experiments in the treatment of disease the allocation of alternate cases to the treated and untreated groups is often satisfactory, is because no conscious or unconscious bias can enter in, as it may in any selection of cases, and because in the long run [italics in original] we can fairly rely upon this random allotment of the patients to equalise in the two groups the distribution of other characteristics that may be important. Between the individuals within each group there will often be wide differences in characteristics, for instance in body-weight and state of health, but with large numbers we can be reasonably sure that the numbers of each type will be equally, or nearly equally, represented in both groups* [247]*."*

Further promotion of the RCT by Cochrane [249] and others in the 1970s laid the cornerstones of the EBM movement [20]. The early EBM protagonists developed a hierarchy (or pyramid) of evidence with systematic reviews on top and case studies at the bottom, [60,164] on the basis that methodological designs at the top of this hierarchy were less prone to bias (and therefore more likely to lead scientists to the truth).

Viewed from the ideal limit concept, bias is depicted negatively as something that distorts the truth (which would otherwise be pure, universal and attainable). Bias has been defined by EBM scholars as "a cause of systematic error" [228], a "deviation in judgement" [140], an inability "to approximate the truth" [227], "a deviation from the truth" [226] and "the lack of internal validity or incorrect assessment of the association between an exposure and an effect in the target population in which the statistic estimated has an expectation that does not equal the true value" [225].

As EBM deliberately sets out to support clinical decision-making in the single case scenario of the clinical encounter [3], cognitive biases refer to failures in clinical reasoning to correctly estimate the ideal limit ('true' prevalence, incidence, risk), inspired partly by the works of Kahneman and Tversky [140,250–252]. These authors explored "the cognitive and the psychophysical determinants of choice in risky and riskless contexts"[73] through typically well controlled studies using economic modelling and gambles. These notions were further developed by others for healthcare and include, for example, anchoring bias (undue emphasis given to a salient feature in

the consultation), ascertainment bias (thinking shaped by prior expectation), search satisficing (having found one diagnosis, others are neglected) and many more [140].

The various markers of methodological quality proposed by EBM scholars – for example, to use the RCT design where appropriate and possible, to ensure adequate sampling and sample size, to analyse data on an intention to treat basis and to publish all findings whether positive, negative or equivocal [23] – are all ultimately designed to reduce bias, and hence help scientists get ever closer to the ideal limits as the facts 'out there'.

7.3 Bias as inter-esse

But bias can be understood very differently from a knowledge-power structure view as developed by Isabelle Stengers. She refers to interest in the original Latin sense: 'inter-esse' meaning 'to be situated between'.

She claims that no scientific proposition can, in any relevant sense, be called 'true' if it has not attracted 'interests'. Knowledge is, according to her, built on attachments, interests, attractions. In 'Power and Inventions' she writes:

> *"However, in most cases, a "fact" is not in and of itself so talkative. Its significance, as well as its recognition, involve a history that is produced by active strategies. Whom will it "interest" – that is, who will agree to associate his [sic] research program with it, to be situated by it, that is, let it "be between" (inter-esse) his own questions and those that produced it? This is a crucial question,*

because what we have to call the creation of a reality depends on it. Indeed, reality is of course not what exists independently of human beings, but that which demonstrates its existence by bringing together a multiplicity of disparate interests and practices [32].*"*

In 'Sciences et Pouvoirs' she states in her introduction (Au nom de la science) that she intends to refute the traditional claim that science and interests are opposites and that science (and associated claims of objectivity and truth) can be obtained only by "purifying" science from subjective interests. On the contrary, facts become facts by attracting subjective interests:

"They deserve this title [of being a fact], not because their existence has been proven by empirical science (since subsequent empirical studies can always destroy it using new technical approaches and sources); but because they have become a crossroads for heterogeneous practices, each with different interests, each of which has required the phenomena in question to be able to relate reliably to their questions and interests. [238]*"*

[Quote translated from French]

Stengers' argument should not be mistaken for a social constructivist claim. She does not argue that facts are mere social constructs. She does not claim that all propositions about reality are equally true and that their status as facts (or non-facts) depends solely on social consensus. Rather, she claims that truths are accepted as facts only when they become interesting. When Pasteur's discovery of lactic acid fermentation became a

major scientific breakthrough, it was not mainly due to the discovery itself or the rigour of his methods, but because he was "able to make the social powers work in his favour"[238]

> *"Of course, Pasteur's genius was to link the question of microorganisms to questions of interest to industrialists, farmers and doctors. Why does beer become bad? Why do our cows die of anthrax? How to fight against epidemics?*[238]*" [Quote translated from French]*

She goes on to argue that there are many truths that never become accepted as facts:

> *"It is impossible to count the number of proposals that "could have" had a chance and become viable, and were rejected with a shrug ... because for a laboratory "result" to be the origin of a process, a device, a product of interest to the life of society, it is necessary that it has changed hands, that it has become "interesting" for a collection of actors other than the "competent colleagues"* [238]*." [Quote translated from French]*

Her argument is that since facts are in the most fundamental sense built on interests, all science is (in this sense) inherently biased. Facts are never absolute but result from an interest, a bias towards a certain line of questioning.

This perspective aligns broadly with that of other feminist philosophers, notably Martha Nussbaum, who have argued that facts are invariably value-laden and, furthermore, that

an actor's emotional and moral position in relation to a 'fact' (reflexively surfaced and examined) strengthens rather than weakens scholarly inquiry [253].

Stengers' perspective also has some resonance with Foucault's important argument that power and knowledge are intimately related. The quest for knowledge makes people 'visible' by subjecting them to observation, quantification and classification. Moreover, this visibility is integrated into people's own vision of themselves: "He who is subjected to a field of visibility, and knows it, assumes responsibility for the constraints of power; he makes them play spontaneously upon himself [254]." Knowledge thus enables governance by governing people's vision and reasoning. This also implies that without discipline, without controlling people's gaze, there would be no knowledge. In fact, Foucault insists on the productive role of power: "We must cease once and for all to describe the effects of power in negative terms: it 'excludes', it 'represses', it 'censors', it 'abstracts', it 'masks', it 'conceals'. In fact, power produces; it produces reality; it produces domains of objects and rituals of truth. The individual and the knowledge that may be gained of him belong to this production [254]."

7.4 Duality of bias challenging the ideal limit

Stengers' argument implies that there are at least two different meanings of the term "bias". When viewed from the perspective of the ideal limit theorem, bias is viewed negatively and unproductively as anything that distorts the comparisons between groups. Thus defined, bias can potentially be eliminated using technical procedures and checklists. But bias can also be defined in terms of a value-driven perspective on what

is worth studying or taking into account. This kind of bias cannot be eliminated. It is unavoidable – and potentially productive and even necessary. Indeed, it could be argued that without bias, there would not be any truths at all. This stance was set out by Thomas Kuhn in his classic text The Structure of Scientific Revolutions, in which he showed that at any given time, scientific inquiry is both shaped and constrained by a particular collective framing of a problem and agreed empirical methods for studying it; only when this agreed approach fails to answer newly emerging questions do some scientists break away from the collective narrative and seek to establish a new paradigm [14].

At first sight the two views on bias described above appear incommensurable. Stengers' perspectival kind of bias might get her to a particular research question and a particular RCT, but this is a different kind of bias from the biases of the clinicians who allocate the sicker patients to the active drug by peeking at the randomization code or of the researchers who decide not to publish the study after all because it produced negative findings. Yet, in several ways, Stengers' perspective teases out and challenges the assumptions that underpin the truth concept dominant in EBM.

Firstly, in line with Stengers' critique of an alleged disinterested science, the ideal limit theory dismisses or overlooks a subjective element, such as experience, will, consciousness, agency or indeed interests. Between the intervention and the outcome, unknown chains of multiple material causes (chemicals, DNA, heat) and human causes (patient choices, social interaction) are lumped together in pursuit of a single intervention-outcome correlation. Patients and clinicians are understood and treated to behave like dices, billiard balls, beings without internal life tending towards ideal limits

as the rules out there, which apparently govern things like behaviour, agency and interests. If RCTs are assumed to be able to find those rules, what reasons are there to believe this? Are human interests just epiphenomenal to a material world [255]?

Secondly, in contrast to Stengers' view, the ideal limit theorem assumes the possibility of full separation between the observer and the human research participant. A position of non-inter-esse is deemed feasible. It is assumed there are facts out there, with the potential to be discovered by disinterested, objective observers (as long as they work hard enough). Measurement is considered to be attainable without disturbing the topic of enquiry and the ideal limit it tends to. Is this indeed how the world works and if so what proofs or reasons do we have to support this? The bias that should be avoided when comparing groups would not necessarily require a disinterest of the observer, but at least an equal interest in both groups, *equipoise*. Yet even the most ardent EBM proponents view this as unrealistic as this would dismiss the role of the two other EBM pillars (clinical experience and patient values) and violating what Sackett called 'trust in the physician–patient relationship'[256] confirming Stengers' point.

Thirdly, Stengers' inter-esse invokes the idea of goal-directedness of scientists feeding back and affecting their activities in real time. We suggest clinicians tend not to think in keeping with statistical causation, but more in terms of Aristotle's causes. Dispositions (an innate power of entities to cause) [204,257] and final causes (purpose defining what things are and can do) are typical in clinical practice [223,258]. Patients are considered to have tendencies (towards getting better or not), and not to behave like dice (either get better or worse and nothing in between). At the same time actions are not left to

abilities and pure chance. Means are pursued to reach certain ends. Equally, a kind of feeding back is assumed in the ideal limit theorem. There is not just the belief that average outcomes in the long run tend to their ideal limit; the ideal limit is used to make predictions for other patients and events that were not in the research study. This supposes that somehow a future limit has some kind of an impact on individual patients today. It is perhaps remarkable that patients, new treatments, even whole complex interventions already tend towards an ideal limit that will only be revealed in the long run [239]. But how could and why could this happen?

Fourthly, the ideal limit is considered to be stable over time and pointing in one direction. This is underpinned by the assumption of a deterministic world view with an eventually predictable future if only we knew all properties and conditions of the world. Current methods of measurement may be failing and human interest blocking the view, yet all events in the past, present and future are fixed already. That reality not only conflicts with current understandings of particle physics that see only probabilities as deterministic, and the occurrence of actual events as fundamentally unpredictable. More importantly, it conflicts with the concept of human free will [255]. In a fully deterministic world, there is no place for human interests. Is reality really like that? Stengers' notion is one that questions a fixed reality by stressing the importance of interaction and agency.

Fifthly, Stengers draws attention to framing as interests shape concepts and ultimately facts. Defining diseases, interventions and outcomes (for example when designing RCTs) is essentially a subjective endeavour. A purification process is needed [21,89] to describe

the specific problem as seen in real patients in one of the many categories as defined by the medical community. The ideal limit comes with the expectation that these in one way or the other correspond to a real world. But why is that acceptable? The ideal limit may help to deny any claims of effect from a certain framing in the long run, but in itself it does not provide any assurance that framings overlap with those in an underlying reality.

7.5 Towards a pluralistic understanding of truth in EBM

The assumptions underlying the ideal limit theorem are largely metaphysical; they can be neither proven nor refuted empirically. The EBM community could very well proclaim that the world is deterministic, that our framings reflect how reality is, that an understanding of inner life or mental causality is unnecessary and that these beliefs are justified as EBM has been so successful.

However, the ideal limit theorem does not suffice for what EBM intended to do [3]. Yes, it provides a truth concept for large groups and frequent events (requiring a belief that similarity can be established unequivocally). But conceptually, it cannot deal with the single case scenario in the clinical encounter – the original situation for which EBM was developed. In that context a patient and a clinician need to overcome the so-called philosophical problem of induction, the inability to predict the future. They will have to make a risky inference; hence, even when the premises are correct, the conclusion and outcome may not be.

As discussed in the chapter on reasoning, because the problem of induction cannot be overcome fully, making inferences is about evasions: finding ways of reasoning to evade the problem and achieve the best prediction possible [28]. In EBM the dominant evasion is the frequency-type evasion. This evasion assumes we can be right in say 95% of the cases in the long run.

Obviously in the clinical encounter this kind of reasoning is problematic: there is no frequency of a single event. It is impossible to say whether a single patient is part of the group who will benefit or part of the one that does not. The ideal limit is about the tendencies of a group as a whole, not of individuals in that group. In practice, other evasions are needed [90] that can embed frequency data in the single case, such as Bayesian (updating beliefs) [28], mechanistic (how does it work) [212,259,260] and again means-ends reasoning (working towards a goal) [223,258]. EBM teachers are keen to educate medical students in Bayesian reasoning because it can be applied to individuals, focusing chiefly on how probabilities of a diagnosis are conditional on certain signs, symptoms and wider determinants [261].

It follows that if (and to the extent that) the goal of EBM is to support reasoning in the clinical encounter, then the ideal limit as the sole concept of truth appears to be both conceptually and empirically insufficient. The cognitive biases observed in highly controlled lab studies as described by Kahneman and Tversky were already contested in the complex reality of clinical practice [140,252,262] which makes attempts to 'de-bias' clinicians [263] futile.

As Norman and Eva [252] write:

> *"Firstly, thinking about cognitive bias solely as a source of error is inconsistent with the psychological literature on the subject. In psychology, heuristics and biases are viewed as efficient mental strategies with which to deal with an uncertain and ambiguous world. On many occasions they work; occasionally they fail. But they are not intrinsically bad. Secondly, although many (but not all) of these biases have been established experimentally, this occurred in situations in which variables could be manipulated singly and clever manipulations were designed to induce error for the sake of determining if the biases exist. Even experimentally it is hard to observe 'biases' in situations where the biases lead to the correct answer. By contrast, when trying to conduct a retrospective examination of a diagnostic error, there is no way to deduce the presence or absence of any error as errors are overlapping and there is no process trace to indicate when and if a particular error arose in the diagnostician's mind."*

This more nuanced and positive perspective on bias resonates with Stengers' notion of 'inter-esse' and raises the question whether there are alternative truth concepts that will help us to differentiate valid inferences from less valid ones while overcoming the challenges mentioned above.

This is critical not just for the clinical encounter. The same is true in other contexts, such as policymaking - perhaps even more so [264]. Value-laden, situated activities, dealing with non-frequent and complex challenges need more than just the ideal limit theorem.

They require “making explicit the premises and values on which each side has built its case [265].”

Philosophers tend to deal with truth in general and not with single case scenarios. As clinicians, however, we deal with the latter. Accordingly, we would like to illustrate how different philosophical conceptions of truth could come in to play using the following case as an example. The case is fictional yet inspired by three actual clinical encounters that I had.

Case

On a Sunday afternoon at an urgent care centre in a UK hospital I see Ms T, who presents with symptoms of a sore throat. She is 25 years old and tells me that it started 2 days before. She may have had a fever, but she is not sure. Drinking is painful. She does not have any other medical conditions, but she does smoke. I ask what her main concerns are and why she has come in today. She tells him that her housemate got antibiotics from her GP 10 days ago; the GP (allegedly) said that it might be scarlet fever. Ms T’s main worry is whether or not she has scarlet fever and whether she needs to start antibiotics. On first impression the patient does not appear to be very unwell. She is not drooling or coughing, and her breathing is normal. I see on the computer system that the triage nurse took her temperature. It is 37.8. She is able to open her mouth on examination and SW sees two enlarged tonsils, with some small white spots. They appear quite symmetrical. Her skin looks a bit red in her neck, but the lighting is not optimal so it’s hard to be certain.

At this point I'm not quite sure how to proceed regarding her question about scarlet fever. I trained and worked as a GP in the Netherlands and is aware of the guidelines both there and in the UK. Based on what I remembers from them, I understand that scarlet fever [266] *is a notifiable condition in the UK for which antibiotics should be considered and swabs taken while in the Dutch guideline notification, swabs or antibiotics are not necessary routinely* [267]*.*

I say I'd like to discuss her case with a colleague in the next room. E is ok with that and I go to the next room where I see DH, another GP. I explain the case to her briefly. She says she's not sure either, but as Ms T appears otherwise well, she thinks scarlet fever is not on top of the differential and she would await events. I agree with this and returns to Ms T. I explain that me and my colleague think that a viral infection is more likely than scarlet fever now and that she can await events without antibiotics as it is likely that the symptoms will settle over the next week. I advise her to see her own GP if things do not improve or come back to the urgent care if she feels more unwell or cannot swallow anymore.

7.6 Analysis

In order to inform the decision whether or not to prescribe antibiotics, I could look at the effectiveness of antibiotics from a Cochrane systematic review of RCTs [268] as reported in the Dutch guideline. This shows that after 7 days of antibiotics use, the symptoms had settled in 87% of the patients, compared to 82% in the placebo group. That finding is however of little help to Ms T as she cannot have the antibiotics 100 times

in parallel (in which case she should recover about 87 times if she would take antibiotics and 13 times not). She will only have antibiotics once, and there is no way to tell beforehand whether she will be part of the 87% who improve on antibiotics or the 13% who do not (similarly, 82% who recover or 18% who do not if she did not take them). Thus, the ideal limit theorem does not function in the single case scenario as a way of predicting the future.

However, I can use the frequentist findings from the Cochrane review to update his a priori belief about what is likely to happen to his patient (hence, taking a Bayesian perspective). Yet, in itself this is still not good enough to draw a conclusion, as there are many other eventualities to consider – such as the likelihood of Ms T developing suppurative or non-suppurative complications, allergic reactions and bacterial resistance. Some of those can be updated with frequentist data of course, but they would leave me with many conflicting and disjointed inferences. To make a decision, I need to incorporate concepts of truth that can deal with the practical problem facing him and his patients in the here and now.

I try to make a kind of inference that is limited by what he is trying to achieve for the near future – namely, Ms T achieving relief of her symptoms or preventing complications. Here we see a kind of goal directness or mean-ends reasoning [223] delimiting what I can infer from all possible inferences. Aristotle referred to these as final causes [269] , purposes pushing back on reality. He also proposed the concept of orexis, in the words of Nussbaum an inclination, desire, reaching out to explain that actions are guided by both goal-directness and the “limits imposed by the world of

nature" [270] What's more, my goal-directedness is a certain kind of goal-directness: it is virtuous, the aim is to do good, instead of bad. This is reflected in work by several authors [233,235,271] who question whether the application of epistemic virtue (or vices), morality, justice and human rights differentiate truth from untruths.

What I can know and infer is further limited by his reasoning, he has no certainty of anything 'out there'. This points to a kind of correspondence concept of truth: (transcendental) idealism as argued by Kant [272], Lonergan [104] and many others. They point out that to make a good inference SW is dependent on reasoning. Yet many less explicit kinds of knowing, such as tacit knowledge and practical knowledge (as captured in mindlines [10,16]), and less rational kinds of thinking (intuitively, fast [273], subconsciously, or using heuristics [6,70] and gut feelings [198]) do not fit coherence, idealism or any other dominant theories of truth very well. Rather, they are situated, individual and oriented to answering the moral question "what is the best thing to do for this patient at this time, given these contingencies along with wider evidence?".

In consulting my colleague, I use consensus (albeit on a very small scale) as a concept of truth to limit all the possible inferences. In this view, it is not just that two GPs know more than one, their agreement actually makes things more true. This is somewhat reflected in the Bolam test, a criterion in English medical law: what a community of clinicians considers a reasonably competent clinician would infer in a specific case is considered true [274]. But consensus between patients and clinicians is often not clearly expressed. A polyphonic concept of truth as developed by Bakhtin, where the combined utterances of participants in a meaningful conversation limit the possible inferences, Is

perhaps better suited. SW is never an objective observer and cannot be disinterested as Stengers points out in the more general case. Throughout the interaction between me and Ms T the concern about scarlet fever emerges, not out of agreement, but through interaction it becomes an entity that needs to be addressed.

Pragmatism, the idea that what works or what is useful is true as developed by William James [275] and others, might fare better in the single case scenario. However, it is not so clear for SW to see what 'works'. Following frequentist data, giving or not giving antibiotics both 'work' in reducing symptoms in 7 days. Equally, to the complication rate is so low (the number needed to treat for rheumatic fever is about 1 in 1.5 million [276]) that it is hard to see how diagnosing or not diagnosing scarlet fever (with the result of notifying or not notifying) in Ms T case achieves anything for her or is useful for anyone else. Philosophers have drawn attention to the point that in practice many actions 'work' and it might be more important to focus on what does not work or has negative consequences (i.e. negative pragmatism [277]).

In this case, I try to adhere broadly to the relevant guideline (as he remembers it) to make the best inference for Ms T, but becomes puzzled as guidelines from two countries conflict. Both of them are based on about the same international systematic reviews and RCTs but through the guideline development process somehow ended up with very different recommendations. A constructivist concept of truth operates here. The Dutch guideline makers chose to put [267] for instance less emphasis on studies from before 1960 since then rheumatic fever (and other non-suppurative complications) has become so rare, that a GP will mostly likely never see it. Moreover, I have to take into account

that a UK trained colleague seeing Ms T the next day has no awareness of the Dutch guidance. As a result, I need to align any inference on scarlet fever with UK practices. Furthermore, as I deviate from what Ms T might find online on scarlet fever, I can explain more carefully my reasoning, bringing in elements of multiple concepts of truth. For instance, that I'm not convinced by the small risk of serious complications or the added effect of antibiotics in the long run (kind of ideal limit via a Bayesian evasion), that my colleague broadly agrees with me (an element of consensus), that I see diagnosing scarlet fever as not useful in helping Ms T achieve her purpose (kind of goal-directed pragmatism) and have noticed other recommendations based on similar evidence (a kind of constructivism).

7.8 Conclusion

In this chapter, I have argued that bias is in a dual, complex, necessary, unproductive as well as a productive conjunction with truth. It presented characteristics of the ideal limit theorem as a dominant theory of truth in EBM and looked at several other theories of truth that operate in the clinical encounter. In the next chapters I present the results from my empirical work through the lens of the issues and concepts discussed in this chapter as well as chapters 4, 5 and 6.

Chapter 8 Mindlines online: the nature of informal clinical knowledge

This chapter is an updated version of a paper co-authored with my supervisors called 'How Knowledge Is Constructed and Exchanged in Virtual Communities of Physicians: Qualitative Study of Mindlines Online' published in the Journal of Medical Internet Research [92].

8.1 Introduction

In the previous chapters I laid out a philosophical critique on the beliefs and fundamentals of the classic EBM paradigm with regard to types of knowledge, styles of reasoning and theories of truth. I argued that evidence-based knowledge creation could be described as modernist practice of purification, offers a philosophically limited perspective on knowledge, focuses narrowly on frequency type reasoning as an evasion of the problem of induction and primarily commits to metaphysical assumptions of the ideal limit theorem.

In this chapter and the next, I present two empirical studies in which I explore how meaningful and valid knowledge is and could be produced in practice, reflecting on aforementioned epistemological criticisms. In this chapter I present the results from the study three virtual networks as set out in Chapter 3.5, to address the question "how is knowledge actually developed and shared in informal communities of clinicians?".

8.2 Results

The activity (rate of posts) of members varied hugely both within and between the three networks we studied. We found the TGG group in the UK to be by far the most active of the three. On the topic of cholesterol and statins, there were more than 1300 posts over 2 years. Most members did not contribute; all discussions were started by 64 individual members and all posts were written by about 150 individual members. These posts varied considerably in length. Most were informal and very short, consisting of a word, an exclamation or a link, though there were interspersed with longer postings that included segments of patient narrative or presentation of an argument. Most discussions had responses, with some threads extending beyond 30 posts. In Eyr and HAweb there were fewer than 100 posts in 35 discussions over the same timeframe. These posts tended to be longer and more formal, and attracted none or a small number of responses.

The TGG group was accessible via a browser, but also via the Facebook app on a smartphone, which made it easily accessible and integrated with other activities and friends on Facebook. Eyr and HAWeb are more dedicated services without a smartphone app. A member had to log in first to see discussion but would get further updates and could reply via e-mail directly, once they had engaged in one.

Many of the GPs who did post on TGG appeared to be relatively young and inexperienced (judging by their profile pictures, postings such as worrying that they may have posed a ‘silly question’ and saying that they were doing locum work). The members on Eyr and HAWeb appeared to be more senior, since their postings featured concerns about issues relevant for practice owners or revealed extensive previous experiences relating to the issue at hand.

All networks had administrators and a protocol on rules and what kinds of postings are permitted, but in practice moderation appeared to be very limited on these networks. No clear signs of interventions by an administrator were found on any of the threads studied (though we did not find any breaches of protocol either).

The shape and nature of knowledge online

- *Guidelines mentioned only rarely.* As Gabbay and Le May's also observed, we found few occasions where GPs referred to evidence and guidance on the TGG group. In a sample of over 1300 posts, there were just 45 links from members to online content elsewhere, 17 posts containing the word guidance, 50 with the word guideline(s), 25 the word evidence and 29 references to the National Institute for Health and Clinical Excellence (NICE, the national guideline developing organisation in the UK). Of 37 links to other websites; four were to NICE, three to GP-Update (provider of specialist courses and online learning for GPs) and four to GP Notebook (a database of clinical medicine topics for GPs with a search facility). In 46 of the 76 discussion threads, the words NICE, evidence, study, guidance or guideline were not mentioned in any post. On Eyr, members' references to guidance or guidelines were similarly sparse, but linking to other papers and sources of information occurred more frequently: 47 times in 26 discussions. On HAWeb there were few links to other websites, but referencing to guidelines happened in 18 out of the 56 posts over 9 discussions.

The higher proportion of posts that referred to guidelines on HAWeb and Eyr appeared to be related to the more formal nature of the posts. Many of these specific references

were presented to the community as useful to solve the problem at hand and could be posted without any further commentary. Other links were posted to draw attention to medical news in the media: to the debate on the usefulness of statins in lowering cardiovascular risk; the influence of vested interests on guidelines; or how statins were discussed in new journal articles.

Strikingly in many cases on TGG and Eyr, members did not state which guidelines or evidence they were referring to. Rather, a 'nominal' perspective towards guidelines was taken ('guidelines' referred to in name only, and in a generic sense). For example:

> *"As the guidelines state – those suspected of familial hypercholesterolaemia should be referred to a specialist for most likely high-dose statins" (Post on TGG)*

> *"I keep getting into patients in nursing homes in the age group 90+ with 80 mg statins in addition to 15 other medications. Separately certainly correct according to guidelines (written by people who are sponsored by industry)." (Post on Eyr)*

On HAWeb this was not the case, as for GPs used the term 'standaard' which is the specific name for guidelines produced by NHG (hence, from a GP-funded agency independent of industry influence). In the Netherlands, there is almost no competing guidance for GPs, which probably explains why these GPs talked about '*the* guideline' rather than 'guidelines in general'.

- *Case based reasoning.* The concept of mindlines envisages knowledge being constructed collaboratively by sharing cases about patients and their situations. In the TGG group, there was a strong tendency to start a discussion or post a question based on a clinical case. In 44 out of the 76 discussion threads, initiators started off with describing a patient they had recently encountered. Statins were not necessarily the main topic on which they sought peer advice, but were mentioned as part of explaining the clinical situation. Group moderators and the GPs themselves made great efforts to protect patient confidentiality. In all cases, personal details were anonymised and adjusted to make re-identification impossible.

There were some references to actual cases in Norwegian Eyr too, but none in the Dutch network. It is not clear what explains these differences, but it may be coincidental and related to the small number of posts about statins and cholesterol, since on other topics in HAWeb members do present patient cases. An alternative explanation is that the prevailing debates among epidemiologists have taken place predominantly in the UK and attracted extensive coverage in the medical and lay press there; some exchanges have been acrimonious and linked to allegations of conflicts of interest and the threat of legal action against the British Medical Journal. Dutch GPs, in contrast, were exposed over the same time frame to a much more consistent message from a single, uncontested national guideline.

- *A focus on 'what' rather than 'why'.* Where clinical cases were shared in the online group, they were mostly presented in a short, clipped, telegraphic style of writing. Elaborate narratives taking a more holistic view of the patient's context, preferences or

social relationships were uncommon. Rather, clinical findings and test results were used to describe patients and frame the problem. For example:

> *"Presenting a woman 51 years. BMI around 28. Little exercise. Former smoker but stopped 5 years ago. Before this 20 pack years. BP average 140/100. Total cholesterol 6.8, LDL 4.1. Other normal. No familial risk of cardiovascular disease." (Post on Eyr)*

The above quote is typical of the style in which a patient was presented "objectively" to peers in the virtual network. There is a strong emphasis on quantitative biomarkers and on aspects of lifestyle, such as smoking and exercise, that have been identified in research studies (and the guidelines that draw on them) as risk factors for cardiovascular outcomes. In this way, the account is strikingly parsimonious, omitting aspects of the narrative that are not directly related to cardiovascular risk factors. There is no detail, for example, about how she articulates *why* she is not exercising (or what could be done about it), what made her stop smoking, her life story, family circumstances or the way she reasons about her own health and lifestyle. Perhaps the reason for this is that the virtual forum is being used to help *interpret* the guidelines on a case-by-case basis; whilst other factors (such as patient preference and circumstances) will also inform the clinical decision, this is not the aspect of practice for which the posting GP is seeking input from others.

- *Emphasis on outliers*. Another finding when considering the presented clinical cases was that these posts were most frequently about unusual situations and outliers. For

instance, out of the 44 clinical cases in the TGG group, 16 covered a case of familial hypercholesterolemia, a condition affecting only 0.2% of the UK population [278]. On these networks, it appears that what becomes explicit, what gets discussed and forms the basis of learning does not reflect the bread-and-butter knowledge of clinical practice but uses unusual cases to extend and challenge that knowledge (see discussion).

- *Material, regulatory and practical constraints on knowledge use.* In all three networks, members frequently referred to regulations, directives, systems and financial restrictions in which decisions were situated. They discussed knowledge in relation to their role in the welfare state, the perceived influence of industry and important socio-economic and ethical considerations. Recommendations on statins often conflicted with economic reality – for instance in this post on problems relating to funding and quality indicators for statins:

> *"Module 2013 states at indicator 11B: preferred means simvastatin alone. So no pravastatin, as is preferred in accordance with the revised standard CVRM 2012 simvastatin. But now we had run into the problem that in practice many patients in the past had side-effects of simvastatin and quit. With provide an inexpensive alternative - restarting pravastatin - you score worse on the new indicator 11B in 2013." (Post on HAWeb)*

Here a GP laments that the indicator 11B recommends that he prescribe simvastatin to lower cholesterol as advised by the national guidelines on cardiovascular risk management (CVRM). But if he wants to offer patients who have side effects a better

alternative in the form of another cholesterol lowering drug (pravastatin), his 'quality score' (according a prescription module from an health insurance company in order for the GP practice to gain the status of a 'plus practice') goes down. This illustrates how knowledge, and decisions based on that knowledge, are embedded in the local context in practice – and also how they are influenced by the 'action at a distance' of well-intentioned national policies to improve prescribing quality but which are insensitive to the granularity of individual cases. In this case, a financial remuneration system competes with knowledge from guidelines and also goes against a solution that would seem more beneficial to the particular patient being considered in the here-and-now.

- *Not-so-explicit knowledge.* The concept of mindlines entails that much of the knowledge shared is tacit, inexplicit and more than can be told [159]. Though the very nature of tacit knowledge makes several authors suggest that this is inherently uncodifiable [279], others note that tacit knowledge can be surfaced and measured, albeit indirectly [280]. If defined as the knowledge subconsciously needed to perform things that are the focus of our attention [182], tacit knowledge can be surfaced through storytelling, modelling, shared practice and other social interaction. Based in this latter understanding, there is a degree of shared knowledge assumed in most posts in our dataset. GPs would causally use concise reasoning in short sentences, which the poster knew did not need to be articulated explicitly. For example, one GP presented a case like this:

"Need LDL before thinking familial HC, def do TFT and Dm check, poss NAFLD given results, think what you done correct, I would just wait re clinic refer before getting results back in case TFT / DM show cause" (Post on TGG)

This goes beyond jargon as an economic means of using language. The phrase "Poss NAFLD given results", assumes not merely that that list members will know that NAFLD is an abbreviation of 'non-alcoholic fatty liver disease', but also that everyone in this community knows that this diagnosis is relevant to the discussion and a reasonable conjecture given the blood test results.

To give a more subtle example, members of our virtual networks mentioned statins and their use frequently in posts and discussions on unrelated topics. Throughout these posts, the role of statins is not immediately evident as the comments on statins are tangential or incidental to the topic being discussed. Members appeared to embed their knowledge about statins across multiple posts in a non-linear, intuitive way. For example, in this posting in a thread about research on gastric bypass (a surgical treatment for obesity), a participant first declares that he is skeptical about the procedure; depicts much of the research on it to be poor quality, conducted by those with vested interests and heavily influenced by publication bias; and relates some stories of colleagues' patients who underwent gastric bypass. He then brings in statins as an example of a therapy that is likely to have very limited impact on a problem whose underlying cause lies elsewhere:

" I am sceptical about this kind of research, especially as long as they are conducted at private clinics and gastro surgical departments. that promote this. It's possible I told this story before, but I take my chance to tell it again. I worked on one of the largest hospitals in the laparoscopic paediatric surgery. A colleague was tasked with the making prospective study of laparoscopic vs. conventional appendectomy. The results proved to be so discouraging that all "material" was thrown in the shredder. Unfortunately, the pharmaceutical industry keeps this going. I find it reassuring that X writes that we know too little about long term effects of such operations and that the Knowledge Centre needs to look into this. Recently received input from a few colleagues about bypass surgery. One patient weighing 125 kg before surgery, and now weighs 170 plus. My point is that we need to add significantly more "weight" to preventive health. It is quite pointless to repair, be it with statins, operations and other means." (Post on Eyr)

Here, the reference to statins could be explained by members of the community as fitting in an argument of alternative interventions to reduce cardiovascular risk, but this remains subtle and implicit. The poster – probably rightly – assumes that his colleagues will understand the context in which he is referring to statins. And thus, the poster makes a small contribution to a wider 'mindline' on the limited efficacy of statins in the face of multiple risk factors and social determinants.

Another example of tacit knowledge transmission was the use of images and figures. Pictures of parts (a tongue, an eye, some skin) of unrecognisable, consenting patients were posted in four discussions in our dataset, accompanied with a brief wording, along with tangential mention of the use of statins. Most of these involved diagnosing

dermatological cases which are highly dependent on pattern recognition. For example, one person posted a picture of an unusual rash, and (in a thread where several others had offered guesses as to the diagnosis) another GP suggested that this may be a vasculitic rash related to an adverse drug reaction, saying he had seen a very similar rash in such circumstances before. Whereas the other posters in this thread were attempting to elucidate the diagnosis rationally from particular features of the rash, this poster was relying on intuitive knowledge that he was unable to (and/or did not find it necessary to) articulate in words.

Another important way tacit knowledge was shared in these online forums was through GPs' postings on what they would do, without citing any formal sources of evidence or explaining their reasoning. On the TGG group, such practical knowledge exchange was common. For example:

> *"Personally, I would choose to reduce alcohol and assess to solve or aid in reducing alcohol while give her a referral for 2-year test". (Post on TGG)*

Similarly, GPs often suggested what might or might not work, with either no justification or a justification (as in the example below) in which some of the rationale was implicit rather than explicit. In the case below, the poster is arguing that the relatively rare (and usually mild) side effect of depression on statins could become life-threatening in patients with a history of severe depression:

“I would not even use statins, without having much harder risk factors than the currently loaded policy. As a psychiatrist, I treated a few patients which were so depressed that they needed hospitalization, shortly after starting with statins. And rhabdomyolysis by individual interactions and other unforeseen circumstances, is simply not to be trifled with.” (Post on Eyr)

At the same time, these statements often reveal not just how patients should be treated, but also personal experience and internalised preferences of the members themselves. For example, in a discussion on a case where patient has a very high cholesterol level, but (taking other risk factors such as weight and blood pressure into account) a moderate risk of cardiovascular events, members discuss what would be the appropriate thing to do. In commenting on his own post, the initiator of the discussion writes:

"If it was me, I'd want a statin. I think that's what my gut says." (Post on TGG)

In this example, the member is expressing his professional disagreement with imposed policy, referring to how he would feel if he were the patient. In some discussions, members even disclosed their own or their family members’ conditions and how they dealt with them, what medication they would take and how they experienced the services they used (as patients). For example, in one thread, a member presented the case of a patient with raised levels of cholesterol and a positive family history of high cholesterol, in which she states that this is her husband. She is concerned that his GP only discussed the matter via telephone and wonders whether it would have been

appropriate for her husband to go to a specialist lipid clinic. This posting, and the responses to it, illustrate how clinicians' professional knowledge could not be easily separated from their personal experience of illness and risk. Rather than distant abstract explicit facts, clinical knowledge is embodied and concrete, directly affecting these clinicians themselves.

- *Collective reinforcement of knowledge.* The online communities served as a collegial fellowship of professional peers where the expected format is for one person to ask for advice and others to share their experiences and reflections. The discussants do not provide final answers, instead they ask additional questions, give comments, suggestions or theories that should be subject to further experiments and testing. Each answer is not inherently right or wrong, but the multitude of answers provides a broader sense of what is going on and the options for taking action. This is the means by which a knowledge base gets constructed to support a decision towards multiple actions appropriate for the situation at hand:

A *17 year old pt has total cholesterol of 7.9 and LDL 5.6. no other risk factors. no family history of heart disease or high lipids, what would you do?*

B *Why was it done then!!*

B *Diet, weight*

C *Send the blood result to the one that ordered it for them to deal with. She/He should know why they requested it!*

D *familial hypercholesterolemia... send to lipid clinic and family too*

E *Bounce straight back to the one who ordered it. They can decide to repeat it instead of basing a lifetime of advice on 1 possibly accidental outcome. Never do a test if (a)you do not want to know the result and/or (b) will not be able to interpret. Unless you are about to go on vacation or quit that job*

(Post on TGG)

As a basic part of the concept of mindlines, the knowledge about statins is collectively reinforced through all kinds of social interactions and situations. In most cases the purpose remains largely clinical, but members also wrote about statins and lipids outside the context of clinical practice (for example in relation to the allegedly unethical behaviour of pharmaceutical companies in promoting overdiagnosis and overtreatment). Medical knowledge about statins was embedded in a wider array of stories, jokes and noteworthy events. For instance, in a TGG group discussion about a cardiologist who advised to replace margarine with butter, a member states that he likes butter for many reasons and posts a picture from an unappetising cartoon figure ripping his shirt open, revealing his torso with the words “grease me up woman!”.

- *Identity and social support.* As in other topic threads, knowledge about statins was built and shared in postings that also served to support social cohesion and interaction or describe an atmosphere. For instance, a member started a discussion to express her joy

at seeing other members of this virtual group at a face to face event. A member, who apparently was not at the face to face event, asks:

> *"So, what was the buzz? Was it statins versus stilettos or vests versus venlafaxine? Xxx" (Post on TGG)*

Other members added to this discussion (talking about cardigans, the need for more gossip and that they should meet up again face to face). GPs showed great affection for the group; throughout the TGG network there were many postings that they enjoyed the interactions and the contact with peers. For instance, one discussion thread began with a member ventilating about an afternoon surgery in which there had been many complex cases. Participants expressed their support in several ways, and then one participant interjected with:

> *"Been a regular here for such a long time... I love this group" (Post on TGG)*

Such socio-emotional contributions appeared very important, especially when members were posting about the stressful and challenging aspects of their job. In addition, these kinds of posts show an understanding of knowledge as relational on many levels: as indivisibly linked to the speaker, intrinsically interwoven with other concepts, as a mean to connect with others, and to form and sustain a community.

8.3 Discussion

Characterising online knowledge and knowledge processes

The three virtual networks in this study support knowledge processes in an unstructured, self-organising way. They are not about translation in a direct sense, but intermediation [155]: "the messy engagement of multiple players with diverse sources of knowledge" where a limited set of rules and nudges are imposed by administrators and software. As concrete software tools, these networks can be understood as knowledge artefacts 'holding' mindlines but at the same time as 'collaboratively created artefacts for knowing' to support the development and sharing of mindlines [34]. This blurred duality of 'knowledge' and 'knowledge development and processing' encapsulated in one single concept is also key to understand what mindlines contribute to the debate on evidence based healthcare. It reminds clinicians, researchers and guideline developers to stop thinking of knowledge as a simple package, ready to be 'implemented' in every day clinical practice. This may be old news in other research fields and traditions (most notably in the social sciences), but not in healthcare where the often-unquestioned love for 'facts' from randomised controlled trials require stronger rebuttal. We would contend that our findings inspired by the concept of mindlines provide good arguments to move beyond current evidence based practice.

First, our findings support the idea of knowledge and knowing as a social practice of a group, consistent with literature on communities of practice and organisational learning [e.g. 6,7]. Members of the networks have a shared interest and purpose (providing healthcare), they interact and learn together (asking questions, contributing to

discussions, showing affection, behaving like a fellowships of colleagues) and they share a repertoire of routines to solve the problems they face (ways of presenting a case, the use of jargon, critiquing research evidence).

Second, our findings show that on these networks, both explicit and not-so-explicit knowledge is present. Non-explicit, tacit and practical knowledge is present on virtual networks, but in different shapes. It can be seen in jargon, images, stories and 'I would do'-statements.

Third, online knowledge on these networks adhered to the concept of mindlines in that it could be characterised as 'knowledge-in context-in practice' [283]. In the example of the prescription module we see how prescription of GPs is accessible to distant others who are monitoring and rewarding / punishing the GPs in ways that put pressure on clinical practice. Like the Foucauldian panopticon, knowledge and knowledge processes form and are embedded in power relations [254]. By explaining the problems they faced when implementing recommendations, posters offered other members useful insights into the regulations, quality frameworks and payments schemes in what Larry May has coined a 'web of commitments' [284]. For instance, a GP's duty is not merely to apply the guideline to a single patient appropriately, but also to maximise health gains from a limited public fund and/or acknowledge the patient's limited means to pay for the treatment. This could be interpreted as contextual reality (social, financial, legal, etc.) 'pushing back' [285] on clinical knowledge and recommendations, complicating them, but also reducing their abstractness and operationalising the knowledge, making it practical and relevant for the task at hand, right there, right then.

Fourth, storytelling and casuistry are evident in our online datasets. The stories shared tended to be case-oriented, focusing on unusual, rare or extreme events. This is not a peculiar characteristic of the online environment. On the contrary, it has been well described and analysed in a philosophical paper addressing the paradox of using extremely rare cases to teach students about more common conditions ("when you hear hoof beats, don't think zebras"[286]). However, it perhaps illustrates that the virtual network is being used in particular to *extend* clinical knowledge.

We observed a more 'telegraphic' format online than the classic narratives shared among clinicians in face to face interactions [287,288].

The cases largely lacked the "biographic and social context of the illness experience" [289] present in everyday clinical practice. A hybrid kind of case-based inductive inference occurred: casuistry [203] and pathology based at the same time, with the main focus on test results and other quasi-quantitative facts about the individual patient (such as whether and how much they smoked). Notwithstanding that this approach may be partly due to the topic (cholesterol as a blood test based risk factor), the format and style of discussion contrasts markedly with the discussions among GPs in a bygone era in 'Balint groups', where the central focus of case presentations was the patient's subjective narrative and the unfolding of the interpersonal GP-patient relationship in an overtly psychodynamic framing [290].

Validity

As discussed in chapter 4, mindlines lack of a theory of validity of knowledge but based on our findings I would suggest that validity in online communities appears to be about

what works as kind of pragmatic reasoning. If we assess the popularity of some of these networks, we assume that users find much value in them. But the knowledge on these networks almost never clearly comes to a conclusion or single recommendation based on a criterion of correspondence (to reality) or coherence (fitting into a web of beliefs). Rather, their value seems to lie in their ability to support a practical decision in the here-and-now.

The clinicians in these networks tend to say what they would do or suggest what (in their view) might work. The personal commitment to these suggestions is stressed: "personally..." or "I would say…". In making these suggestions and asking additional questions, it appears they do not mind exploring issues of uncertainty. Indeed, by attending to these uncertainties, they seem to gain a kind of knowledge that is useful for them. Uncertainty itself "is not a regrettable and unavoidable aspect of decision making but a productive component of clinical reasoning" [104] .

This relates to Weick's concept of 'collective sense-making' that is embedded in the theory on mindlines. Through discussion, we make things comprehensible in the best way we can collectively, because an ultimate truth or reality remains uncertain [291]. As the ultimate truth cannot be known, discussants are 'scratching around' to make the world as understandable as they can, and only they can judge whether these understandings are useful [291]. As Wittgenstein observed, some problems are readily solved by the addition of data; others (especially those for which a simple answer is impossible) are solved by a *deepening of understanding* [169].

By saying what they would do personally, members appear to adhere to a criterion for valid knowledge similar to pragmatic theories of truth as defined by James: “Ideas … become true just in so far as they help us to get into satisfactory relations with other parts of our experience” [292]. But not fully. Our results show that many views do not come with a clear function, and can be contradictory, disjointed, on a tangent or unclearly related.

An alternative way to look at validity in online virtual networks with a lack of final conclusions would be to use Michail Bakhtin’s concept of polyphonic or unified truth as mentioned in chapter 4 [168]. “Truth” is better conceptualised not as a single thought or post but as the sum total of interactions of posted perspectives, including areas of dissonance and disagreement. This is more than just the summation of those posts; it is an emergent property.

Bakhtin also helps to cope with another problem of mindlines as presented on these online networks. Most network members do not contribute any posts. Only a minority of members does. In how far do the posts represent the ‘true’ mindlines of the collective? This problem is not just limited to mindlines or online networks, but to all notions of ‘collective’ knowledge. In Bakhtin’s view on truth, what is not shared degenerates. Ideas can only thrive and become truthful if they engage in dialogic relationships with other ideas.

8.4 Conclusions

The findings in this chapter are consistent with a definition of online clinical knowledge as mindlines, seeing them as instruments produced collectively by clinicians to base their decisions on; a general concept for both explicit and tacit knowing shared among social groups and reinforcing norms of good practice in a fluid, dynamic and constantly evolving way.

The far less structured interactions on the virtual social networks in this research represent a broad understanding of knowledge associated with many important knowledge theories that may be of use for the guideline community to address their challenges as discussed in Chapter 6. I will consider limitations of this research and the potential consequences of the findings presented here in Chapter 10. In the next chapter, I first present the findings from the ethnography of guideline panels uncovering the beliefs underpinning formal processes of evidence-based knowledge production.

Chapter 9: An ethnographic study of guideline development panels in three countries

This chapter is an updated version of a paper co-authored with my supervisors called 'Can better clinical practice guidelines help reduce epistemic uncertainty? An ethnographic study of guideline development panels in three countries' [293]*, submitted to Academic Medicine in 2019.*

9.1 Introduction

In the previous chapters I proposed that clinical practice guidelines, even when robustly 'evidence-based', could incorporate – to a greater or lesser extent depending on their provenance – multiple different kinds of knowledge, ways of reasoning and theories of truth. In Chapter 8 I contended that knowledge in communities of clinicians could be described in terms of mindlines. In this chapter, I turn to guidelines and present the results from the ethnographic study as described in Chapter 3.5, that looked at how guidelines panels 1) combine (or why do they find that they are unable to combine) different kinds of evidence and 2) how their work was influenced by different perspectives on epistemic hierarchies and approaches to deal with epistemic uncertainty [102].

9.2 Results

The panels that were observed varied greatly in set up and size. NICE panels (UK) were large (up to 25 people) and multidisciplinary (consisting of NICE staff, experts from several disciplines including health economics, and lay members). In contrast, NHG

panels (Netherlands) were usually smaller and consisted of only GPs, although there was also one larger multidisciplinary panel. In Norway the observed panel meetings were multidisciplinary.

In order to give a more in-depth view on the collected data, I have attached three examples at the end of this chapter showing part-fictionalized summaries typical of panels' guideline development practices in UK, Netherlands and Norway respectively. It has to be noted that the NICE panel in example 1 was part of a NICE pilot project that operated in a significantly more protocolized way compared to the majority of panels at NICE.

Different kinds of knowledge

Research evidence was frequently mentioned in guideline development panels, unlike Gabbay and Le May's findings in clinical practice who found that direct reference to guidelines or research evidence was unusual [10]. The overall emphasis was on randomised trials and data from research papers. At NICE these would most often be selected on the basis of review protocols. However, in all countries panels also used other ways to get more literature. Members could be assigned to ask certain external experts for literature on a specific topic, do some snowballing based on literature already discussed, look up other literature online during meetings or even bring in personal work:

> *"I have taken this book [shows it to his fellow panellists]. Yes, I have co-authored it myself." - NHG panel member*

Much like Gabbay and Le May's observations of clinicians though, guideline panel members did not usually read the actual research papers. At NICE, they would work with extracted data from research papers provided via email in advance of the meetings or on handouts during the meetings in the form of overviews, GRADE tables, forest plots and other graphs produced by staff members. But extracted data could also be provided by a dedicated panel member who would be assigned to research a sub question, external experts who would be brought in to inform on a specific topic, or non-panel members from other institutions. At NICE the latter included for instance systematic reviews produced by the Cochrane collaboration. At NHG, GRADE tables could be produced by a separate team within the organisation, as there was no budget to give every guideline panel a dedicated staff member to do this. Example three at the end of this chapter sees a panel also at NHG using a critically appraised topic (CAT) produced by GP trainees.

Guideline panels would discuss papers that were excluded, especially when only a few papers remained to be included, or major studies could not be included based on the inclusion criteria in the research protocol. This could happen for instance because the follow-up duration of the study did not cohere with the research protocol, which would have been established without prior knowledge of the evidence. In these situations, panel members would put pressure to revise the review protocol to include those studies. This was sometimes successful.

Gabbay and Le May found that mindlines would be formed from an amalgamation of all kinds of knowledge such as personal memories and experience, what experts had told

them, some guidance, a lot of tacit and practical knowledge, and other diffuse sources. Knowledge would be shared in many ways, for instance through discussions about individual cases. In all panels we observed instances of all these kinds of knowledge suggested by Gabbay an Le May too, but usually in the context of discussing and challenging the findings from the systematic reviews or singular studies. Members would address contextual issues such matters of litigation, limited resources and cultural differences to judge the applicability of the findings:

> *"We don't have good understanding of this. In Scandinavia people are discharged from hospital to nursing home with more resources than in the UK." NICE Panel member*

There were only a few instances where personal or individual cases were brought in and discussed in detail, apart from the lay persons case at NICE panels. More often panel members would share their professional experience to explain what is currently been done, what could work in practice or how they would deal with the problem.

> *"What we would use is [drug X]. I have seen people on [drug X] tablets, that appears to work" - NICE panel member*

> *"If I do not succeed, I will [try to place an IUD] a second time, and then I will do it with misoprostol and then I usually succeed." - NHG panel member*

These quotes also serve as examples of how panellists would simply mention what they would do in a certain situation without providing a reference to specific evidence to support their claims. This appears similar to what was seen in the virtual communities of clinicians in Chapter 8, yet somewhat removed from the "evidence of the clinic" as in current standards of practice and care pathways evidence, suggested by Knaapen [110].

That some knowledge would not be deemed acceptable for a guideline was also suggested in the ethnography presented here. Some knowledge discussed at the panel meetings was not included in the minutes, linking evidence to recommendations tables, the draft update or the final guideline. The final guideline would contain less information on the panel's discussions than the original minutes. Particularly sensitive issues, such as suspected financial (instead of clinical) motives to do certain operations and procedures, appeared to be omitted.

However, although the focus of most guideline panels was on handling and judging explicit knowledge while observing clear rules and transparency, the findings indicated that much more kinds of knowledge were required to make recommendations. For instance, in this example panel members discuss how to advise on whether to operate a patient:

Panel member 1: *"It's about transparency of that decision making, it's inherent judgement, implicit. We should put something about that in the guideline."*

Panel member 2: *"Can I challenge you on that? It's a bed-end decision."*

Panel member 1: *"Is that implicit, is that transparent?"*

Panel member 2: *"I would have some idea if someone smells of alcohol."*

Panel member 3: *"There is no clear evidence about what is driving that."*

Chair: *"But we can make that statement, that we're concerned."*

NICE guideline panel discussion

Here, the first speaker wanted recognition for the role of implicit knowledge in clinical decision making even in the light of evidence. The second panel member tried to explicate this implicit judgment through contextual knowledge, such as the smell of a patient. This kind of knowledge is typically not included or considered in RCTs. The chair then suggested putting a statement in the documentation of the guideline apparently trying to leave room in the guidance for tacit knowledge.

Observations of nuanced actions by guideline panel members also suggested awareness of the influence of social group processes on knowledge creation. For example, a chair would adjust table settings in order to support certain members to speak up and bring knowledge in or to oppose a dominant member. It was interesting to note that dominant members would not necessarily be a topic expert member. Furthermore, throughout the succession of the clinical update panel meetings, members appeared to become socialised in what was expected of them. They seemed more experienced in the processes of guideline making and acted more in ways that were expected from them by the guideline developing institution, such as not bringing in alternative research data after the research protocol had been established. Panellists would ask more questions

that would be of benefit later in the process (such as explicating the outcomes in the PICO (a model for clinical questions: People, Intervention, Comparison, Outcome) before looking at the data) and gained more routine in creating recommendations.

Members gained practical knowledge through training on methods used in guideline development (such as Bayesian network analysis and cost analysis) and how to decide on crafting a new recommendation (such as when to update or when not to update them). A NICE panel was observed being instructed how to word recommendations, apparently in an effort to remove ambiguity and uncertainty as they were told that NICE gets 500 calls a month from clinicians asking for clarity. The instructor drew attention to research that suggests that people prefer clear English and ensure discoverability. Therefore 'front loading' of recommendations from the committee was required meaning that the messages needed to be short: 15 – 20 words, and no more than 30.

Although formally having a more 'fact-based' approach, there was subtle recognition at NICE of the value of other forms of evidence and guideline and excluding other kinds of knowledge as impossible, or even unwanted. For instance, a NICE Commissioning Guidelines manager discussed the recent 'controversial' asthma guidelines, which were not well received by all clinical groups:

> *"We draw upon robust evidence in the meetings, but what physicians [externally] don't realize is that at the meetings, we do draw on a wider range of knowledge. They just see the research evidence we draw upon. But actually, we are confident*

that our recommendations are sound by drawing on clinical as well as research knowledge."

- NICE Commissioning Guidelines manager in panel meeting

Combining evidence

The findings of the study presented here indicated that combining evidence to make recommendations is not accomplished through consensus. This was markedly clear in the clinical update panels from the pilot at NICE who had been made aware that recommendations could not be based on consensus. The chair often reminded members not to bring in other knowledge (from experience or research not meeting the inclusion criteria, such as observational studies) in order to adhere to the analysis as defined in the research protocol. In practice this proved to be quite a challenge. For instance, a panel struggled with insufficient evidence on outcomes for operating people with cognitive impairment. As they are not allowed to make recommendations based on consensus, they look at tendencies in the provided studies instead:

Panel member 1: *"Do we have enough evidence to keep cognitive impairment as a factor?"*

Panel member 2: *"They weren't in the [X] study."*

Chair: *"We're not allowed to make consensus statements, but the studies showed more positive results for operations in the cognitive impaired."*

Discussion at a NICE guideline panel

The ethnographic observations in this study revealed a tension between guideline development as a technical versus a deliberative process. All panels appeared strikingly text focused. Recommendations were constructed on the spot *en groupe* and beamed on screen for all to see. Excerpt three at the end of this chapter sees a panellist reading guidance out aloud. The wordings of recommendations would be guided by previous recommendations as most panels are now involved in updating guidelines. In contrast to NICE policy, NHG recommendations would be developed to cohere with other guidelines, such as European guidance, guidance from other national medical professional bodies and NICE:

Panel member 1: *"We say it does not make sense ... is completely contrarian"*

Panel member 2: *"There is a new thrombosis guideline from the [Association of Gynaecologists]"*

Panel member 3: *"We can work with their questions..."*

Discussion at a NHG Guideline panel

In this sense, guideline making is a very different (rational, explicit, predominantly technical) process from the development of mindlines (subjective, emergent, predominantly deliberative) [10]. In contrast to the appearance of integration of different voices and knowledge in the published guidelines, deliberative (and non-analytical) processes combing different kinds of knowledge were observed in practice. A multitude of more ambiguous remarks by individual panel members appeared to

significantly shape the discussions to understand the findings from the systematic reviews, chiming with Bakhtin's dialogism (see Chapter 4 and eight)[168]. These voices however generally did not end up in the final recommendations or Leading evidence to recommendations (LETR) sections. LETR sections would present summarised deliberations on how to interpret the findings as concise arguments from a seemingly united panel. In a sense this was the case: panel members were given the opportunity after the meetings to amend the sections that portrayed the group discussions. Panellists would need to confirm their agreement with the final guideline as a whole, including those reflections. A NICE lay member however explained that due to the large size of the update reports, with a lot of medical jargon and given only a limited time frame, at least for lay members it was uncertain if they had completely understood with what was agreed upon in the guideline.

Furthermore, deliberations could not be clearly separated from group interactions outside panel meetings. For instance, in excerpt two at the end of this chapter, panel members made use of the informal time before the meeting to discuss relevant cases before the meeting officially began. During the meeting, they drew on personal experiences to challenge the data presented and make alternative recommendations. In this way issues and context emerged organically and unstructured. Yet, attempts were made to capture deliberations and decision making in a structured way. For instance, when RCTs or observational studies was not available, one panel turned to the Delphi method, another to the DECIDE tool. These tools involve a kind of prioritising, valuing certain outcomes over others on prespecified domains such as acceptability, feasibility and resource use [294]. Panel members would receive training for these tools.

Still, a successful application of these tools was not observed in the observed panels. For instance, when deciding on a clinical decision tool, panel members blamed the DECIDE tool advisers to push the group in a certain direction. Also, panel members tried to change the initial valuations because they did not like the conclusions.

Different perspectives on epistemic hierarchies and approaches to uncertainty

Of the three examples, the NICE Clinical Guidelines Update panel (Example 1) most closely espoused the goal (in the words of a senior member) "to eliminate mindlines from that translation". It has to be stressed again that these panels were a pilot project within NICE and significantly different from standard panels at NICE. These panels consisted of a NICE technical team, topic experts, lay people and non-expert standing committee members, working to make rapid updates of parts of existing guidelines. The standing committee members (consisting of clinicians, methodologists, lay people and commissioners) without topic expertise were there to counterbalance the potential dominance of the contributions of the topic experts over the NICE team (who were viewed as less biased). A NICE staff member said *"Mixing specialists with neutrals, reduces evangelistic, a religious belief. It can lead to standoff but that's better than using evidence [selectively] to support beliefs."*

The panel meetings at NICE started every time with checking again declarations of interest. This is consistent with what EBM pioneers deem the positivist position on knowledge in EBM (Djulbegovic and Guyatt, 2009), which holds that there exists an unbiased "view from nowhere" which the guideline panel should aspire to reflect.

Consensus recommendations were not allowed, since the panels were intended to be "fact focused". Protocols usually limited the permissible study types to RCTs (and, occasionally, observational cohort studies) – a measure that was supported by panellists, who quickly learnt not to introduce non-RCT evidence or suggest consensus ways of resolving disputes.

> *"We only talk about research evidence, not other forms of evidence. That is the function of these groups"* - NICE panel member

In the panel meeting described in Example 1, most of the discussion centres on frequentist reasoning – for example on whether the trial samples were representative of knee fracture patients in the UK; the significance of inadequate blinding; the inconsistency of outcome measures, making statistical summation impossible; and inappropriate follow-up periods. For all these reasons, the panel systematically "downgraded" trials (increasing epistemic uncertainty) using GRADE criteria.

Frequency type reasoning also requires a stable concept to model as a dice. The observations in this study suggest that the panel's methods were optimal to achieve a stable understanding of the topic. As we see in Example 1, the topic was briefly presented in a short briefing paper drafted by a (non-clinical) NICE team member. Topic experts were then expected to provide further clarification which proved essential before proceeding. For some meetings of clinical guideline update panels, a conference call with experts preceded the panel meetings to decide on the review protocols (though even when this occurred, there would still be ample discussions on the scope of the

updating). The assumption that a purely methodological assessment process based on PICO questions (even with numerous added items) could lead the NICE team to the assumed external truth was not borne out in practice. Rather, as illustrated by the discussion on how a displaced fracture should be defined, debates on how to understand the topic of interest were frequent and ongoing, even in final stages of making recommendations. This fits with the epistemological lens as the challenge of purification. Defining, categorizing, objectifying, framing, demarcating patients, events and diseases is difficult – and ultimately subjective. Panel members noted that guidelines were only about part of clinical reality. In a discussion on objective criteria to decide whether to operate patients, a panel member remarks that awareness of the specific situation is always required and can never be anticipated fully.

> *"That fingerspitzengefuhll .. there are always things that do not apply to that one patient." NICE panel member*

During the meeting, various kinds of non-frequentist reasoning were expressed, including 'non-analytical' reasoning ("don't like it", therefore the guideline must be "wrong" or "biased") and precautionary principle (an anecdote of a failed procedure in a panellist's relative). The main 'evidence to recommendations' table linked to this guideline stated that "certainty around the majority of outcomes was low …. [due to] … several factors based on a full GRADE assessment". The table thus depicted the recommendation as stemming entirely from frequentist reasoning. It did not capture (indeed, it appeared to have deliberately ignored) the additional reasoning drawn from panel members' experience and intuition.

In the pelvic inflammatory disease (PID) example (Example 2, Dutch NHG panel), a greater degree of epistemological pluralism is evident. Like the NICE knee fracture panel, they begin by using frequentist assumptions to challenge the RCT evidence on when best to remove an intra-uterine device (IUD) in a woman with PID. They are concerned, for example, about a lack of a clear definition of PID (meaning that, in Latour's terminology, the population has been inadequately "purified"); they doubt the representativeness of the trial samples to a primary care population; and the CAT (critically appraised topic) technique fails to produce a definitive answer to the PICO question. In the face of limited frequentist evidence, the all-GP panel quickly reverts to other ways of reasoning. For instance, they use mechanistic reasoning when a panellist suggests that an IUD should be seen as a foreign body that needs to be removed. They also start to ask what would or would not work in practice. This could be seen in light of a repertoire of usability, but through an epistemological lens also as applying a pragmatist theory of truth: the recommendation is valid and true if it succeeds in practical application.

The panel concluded that, given the high level of uncertainty around the trial evidence, an expectant approach ("authority-based evidence" as captured in a different national guideline from gynaecologists) is "probably best". After discussion, they decide to convey the uncertainty in the guideline and recommend the use of clinical judgement on a case-by-case basis. One member proposes that in every case the GP should set out his or her reasons for removing the IUD.

In the example of a screening tool (Example 3, the Norwegian Directorate of Health), the project manager proposes the construction of an epistemologically diverse guideline, reminding the panel that they will be considering "research-based, experience-based and user-based" evidence. Once again, the guideline development process seems to have begun (in previous meetings) with evidence prepared according to frequentist assumptions: a rational assessment of published research evidence using structured tools designed to assess their methodological quality and external validity. The original plan was to use such methods to surface the single, "best" screening tool to be used universally. But in the current meeting of the panel, it appears that notwithstanding the (contested) frequentist evidence, there seem to be pragmatic reasons why different tools may be more acceptable and suitable in different contexts. Panel members begin to stray from the systematic application of a decision tool and instead begin to draw on their own experience (e.g. "what I feel intuitively", based on widespread custom and practice of the use of one tool in Norwegian hospitals, though it is also noted that this tool is not in widespread use in nursing home and home care settings). The palpable sense of frustration as the meeting ends partly reflects the epistemological clash between the frequentists' continuing conviction that a single, "best" tool should be identified and recommended in the guideline and the pragmatists' conviction that recommendations should be based on what will work in practice.

Reflecting on the contrasting views on kinds of knowledge, reasoning and truths at play in the guideline process, a guideline maker observed lots of internal debate on the many different guideline committees' approaches all being practiced in the same institution dependent on the topic at hand:

"First there were the easier questions, there were the RCTs, that's where everybody started. They're easier to understand, because of things like CONSORT. They're easier to make decisions, because you got rules. Before GRADE you had statistics, that we all understood, we all thought we understood p-value, we understood what effectiveness was. But all the time guidelines have tackled harder questions, they moved into public and social care. I think the pendulum swung in order to answer those harder questions; you have to use different types of study design. Now the reproducibility of RCTs is more of a hot item and it swings back [to a purer epistemology]."

- NICE guideline developer

9.3 Discussion

Social scientists who have analysed the practices of guideline development [106,108,109,295] have repeatedly shown how knowledge other than systematic reviews, controlled trials and other designs in the evidence hierarchy is unavoidably incorporated into guidelines. These knowledges are used not just to challenge the evidence (e.g. biological mechanisms and personal experience), but also to act as justifications (e.g. ethical principles and practice standards) for recommendations.

Researchers who have studied guideline development panels ethnographically have observed that evidence is often perceived to be lacking. Indeed, based on one such

ethnographic study of guideline development panels in Canada and the Netherlands conducted between 2007 and 2011, Knaapen et al. identified four kinds of evidence that might be (though which were not necessarily) accepted by panellists as scientific and eligible for incorporating into the guideline: clinical trial findings; biological and pathophysiological principles; ethical issues relating to the research process (e.g. whether informed consent was obtained from trial participants); and "knowledge of the clinic" (i.e. current standards of practice and care pathways) [295]. However, the findings indicated that much more implicit knowledge beyond Knaapen's types of evidence were required to make recommendations, including an awareness of the subjective nature of decision making and the importance of contextual knowledge.

The finding that the guideline production could not be reduced to an effort to achieve consensus among clinical professionals is consistent with another ethnographic study of development of a French cancer guideline [110] by Knaapen who suggested that it was best viewed as the active production of a text – and specifically:

> *"…the emergence and stabilization of a new configuration of biomedical knowledge and practices grounded in the establishment of mutually constitutive links between two processes: on the one hand, the re-formatting of clinical trials into a device for producing carefully monitored evidence statements targeting specific populations and clinical indications and, on the other hand, the increasingly pervasive role of regulatory processes".*[110]

Moreira's four repertoires of judgement [106], science, politics, process and practice as discussed in Chapter 3.5 were evident in the work of all panels, though they played out somewhat differently in each case. The three case examples illustrated how contrasting philosophies of science underpinned many of the disagreements among panel members, influenced the practice of guideline development and, in some cases, became captured in the guideline itself. Whilst multiple epistemological assumptions were evident in each example, the balance of how these were dealt with varied between the different panels.

Knaapen et al. proposed that EBM is less about the quantity or quality of evidence than about "the specific management of the absence of evidence, requiring a transparently reported process of evidence searching, selection and presentation. [..] The legitimacy of Evidence-Based Medicine relies neither on experts nor numbers, but on distinct procedures for handling (non-)Evidence, reflecting its 'regulatory objectivity'"[295].

The findings from this study support Knaapen's conclusion. Yet, it is not just about transparency. The novel focus on the philosophy of knowledge illustrates how at least two contrasting, yet linked, epistemologies are performed in guideline development. This study observed guideline panel setups and processes that set the stage for a "pure" kind of EBM, adhering to a certain theory of truth (the ideal limit theorem – the assumption that there is a hard reality out there which scientific inquiry can get closer and closer to as techniques get more accurate and studies are replicated [91]), a certain way of reasoning ('frequentist', based on the chance that something will occur if

measurements are repeated), using a certain type of knowledge (only explicit data of frequent events) from a standardized form of purification (as in a PICO). But this study also revealed a practice that aims for a pluralist interpretation of EBM, applying multiple kinds of knowledge, types of reasoning and concepts of truth to help clinicians and patients make valid inductive inferences. Guideline panels do not generally surface or discuss their epistemological perspectives, but they appear to practice a mixture of these two positions, never satisfying either fully.

9.4 Conclusion

In this chapter I presented the findings from the second part of my primary research. Most notably, it reveals a tension between a pluralist and purist understanding of EBM in guideline production. In the next chapter I discuss what this means for the tension between guidelines and mindlines and the EBM paradigm.

Example 1: Management of [joint X] fractures (NICE panel)

A NICE Clinical Guideline Update Team (CGUT) panel has 3 meetings planned over 5 months to update the guideline on [joint X] fractures. This is the first meeting. A, the chair opens the meeting at 1000h. In the room are 10 standing committee members, 6 topic expert members and 6 staff members from NICE. A goes over the agenda and asks everyone to update their declarations of interest. Then B, a NICE staff member, presents the briefing paper, while some of us, standing committee members, ask questions. There are two review questions, the first will be addressed in this meeting: What is the clinical and cost effectiveness of internal fixation compared with hemiarthroplasty compared with total [joint X] replacement in people undergoing repair for a displaced intra-capsular [joint X] fracture?

B shows that compliance with this NICE guideline is low at 30%. C, a surgeon and topic expert member, suggests that either the recommendations are wrong, or there is bias in the data. D, another orthopaedic expert member, says that across the board, clinicians do not like it. E (another clinical expert) says that expensive procedures tend to go to expensive people. F, a nurse practitioner, adds that age plays also a part in it. E says that the population in the research was different: in reality the patients are older, in 30% of cases demented, living alone at home and do not generally consent to research. Another factor is that it matters whether there is a surgeon around to do the internal fixation operation. Many hospitals don't have such a surgeon available 24/7. C joins in the discussion "in the old days everyone did everything, but now for reasons of concern for medicolegal reasons. Many surgeons shy away from it, although they are technically able to do it."

The discussion then turns to what an undisplaced [joint X] fracture is. H says that if you put up an x ray, you will get three different diagnoses. E adds: "You back-diagnose: you do a procedure and explain the x ray afterwards". [joint X] fractures are typically

diagnosed based on a one-view x-ray. Lateral x rays are often missing. if you have painful [joint X], you get an MRI scan. J states that classification is historical and rubbish. "What is displaced? Undisplaced is undisplaced in two plains. So we have to nail that in the recommendation."

Then B presents the search flow diagram. He discusses whether anything excluded should have been included. He then goes through the GRADE tables. Studies have been downgraded as the study population are the healthier side of the spectrum.

Expert member C points out that blinding is a joke in this kind of study. G, methodologist from the standing committee, asks if it is possible that some undisplaced fractures ended up in these trials on displaced fractures. E responds: "research will be more precise than the real world."

A notes that there are now 14 RCTs, while there were only 6 for the previous guideline. The panel discusses the outcomes, presented in multiple graphs and tables on handouts. These are quite complex as there are many outcomes to consider. One member would rather have had stoplight graphs. These use colour, like green, yellow and red to indicate whether there is evidence for an effect or not for each outcome-intervention pair.

Several members note that over the years, the procedure and aftercare has changed: there are newer materials, and patients mobilise earlier. They also note that in the study the mean length of stay is 12 days, while in reality it is 19, indicating that the research was looking at a fitter population.

The panel discusses the evidence statements. These are short texts to summarise the findings. H suggests taking the research on length of stay out as Norway (where some RCTs were based) is not representative of the situation in the UK.

After tea, H, the lay member and by background an academic engineer, gets time to tell his experience of his mum who had a fractured [joint X] and how the surgeon had difficulties putting the nail in. The panel listens carefully and one of the panel members

asks whether there was any shared decision-making. E explains that it was presented as 'this is the way forward' and perhaps would have like a more informed discussion.

B goes on the explain the plots and starts with the hemiarthroplasty versus total [joint X] replacement results. E says he does not like the 5-year follow-up study that was included. He feels it does not make sense.
A: But it is the one that set the meeting / review off!
E: I don't believe it.
C: It's not plausible. There is no good reason why difference should be assessed after 5 years.
E: Two-thirds of people are dead at 5 years.
A: It was highly powered, big study!
E: The study was not set up for this outcome. I think they were biased.
C: All patients died from something else.
A: OK, as long as it is all said and captured in the minutes then I am happy with it.

The panel goes on to discuss several of the outcomes. They note quite a big difference regarding whether patients returned to their place of residence:

A: This study shows 53 from 55 come back to where they were. That is much higher than in UK, where it was 50%.
E: But this is Sweden, but that could be a care home. There are many care homes there, but fewer in the UK.
I, a consultant nurse from the standing committee: No, they were not institutionalised, but independently living.

The members continue to discuss in the outcomes of the review. After lunch the panel goes on to discuss the economic review model. The meeting finishes around 1600h.

Example 2: Management of pelvic inflammatory disease (NHG panel)

A panel convenes at 1900h to work again on an updated guideline on pelvic inflammatory disease (PID). All 8 members are GPs. As they arrive, there is fruit, some drinks and sandwiches. They already talk about some PID cases before the meeting starts. After introduction of the researcher (SW) as observer (and without further declarations of interest), they start by discussing a prepared document on PID in women with an intra-uterine device (IUD). It consists of the current guideline text on PID (i.e. the recommendation and a discussion of the evidence that supports it), a new literature search and concept text and a CAT (critically appraised topic – a structured literature review format used widely in the EBM community) done by a couple of GP trainees. The chair directs the discussion to the CAT. The panel members note that there is little literature on this topic and, like the panel, the trainees struggled with the issue that no clear definition of PID exists. In the face of minimal and dated (1970s and 80s) evidence, the new concept recommendation suggests that an IUD does not have to be removed immediately if PID occurs, only if symptoms persists beyond 72 hours. It mentions that this is in line with the national guidance from gynaecologists (NVOG). The CAT is based on the same evidence, yet the GP trainees came to a different conclusion and suggest removing the IUD. They justify this on the grounds that the most recent RCT, dated 2003 and from Turkey, showed that women might recover sooner if the IUD is removed immediately. They consider this trial more relevant as it was the only one set in primary care and had more relevant outcomes. In contrast to the new concept text, the CAT used the GRADE methodology, which is included in the document. A panel member mentions that it was copied from the NVOG guidance.

The panel discusses the issue and one of the panellist's searches for the NVOG guidance on her laptop and reads it out aloud. F brings in that removing an IUD brings costs and raises the need for alternative contraception.

B: "We cannot attach too much value to research studies. [Author's name] is best. It too shows that the benefit of removing the utility IUD is not clear .. there is nothing against it.. to let it in place. Removing after 3 days if needs be, is probably best. This is mainly authority-based evidence."

C suggests letting patients decide, but B thinks that the patient may bounce the question back. F questions whether once the barrier is broken [F refers to the event that infectious agents pass the cervix and infiltrate the pelvic organs, causing the PID], it helps to remove the IUD. She asks if anyone has experienced that in practice. B tells that she had a case where she made a shared decision with the patient to remove it. E responds that her population is just students ".. who want to leave the IUD in because it still costs 150 euro" to have a new one fitted. B remarks that it seems logical to see the IUD as a foreign body.

C reads more of the new text out aloud and after some further discussion F concludes: "There is therefore insufficient evidence for one or for the other. Consideration in general practice is therefore different. Discomfort if you have to replace, chance of pregnancy which you have to do something with, also say that you have to consult with the patient. NVOG says pragmatically to remove after 2 days if not better .. but that is also not based on research. [..] We're going to make a number of choices. You do not have to remove

it, because there is no proof for that. And you can only remove it after antibiotics. That is annoying for the patient, first getting antibiotics from the pharmacy, and then coming back to remove the IUD.

The panel decides to put in the recommendation's note that there is insufficient evidence.

B: We can say that, to the patient, that we do not know. There is equipoise.

G: But what do we put in the text?"

C: As a GP you should know if you take [the IUD] out, and for what reasons.

Then the panel goes on the discuss other topics on the agenda, which include medical treatment for PID and an article on mycoplasma genitialium as an overlooked sexually transmitted pathogen in women. They appear to work collaboratively, routinely and in good spirits. They go on to plan the next meeting. The meeting finishes at 2100h.

Example 3: A screening tool for [patient condition X] (The Norwegian Directorate of Health panel)

A guideline panel at the Directorate of Health has another meeting in order to choose a single screening tool for [patient condition X] to be used in the whole of Norway. The tools consist of short scoring lists, with a certain amount of points given for each answer. [Screening tool A] and [Screening tool B] are both commonly used in Norway. They differ slightly. The meeting is supposed to be structured as a DECIDE process. (DECIDE is a set of tools developed to help rationalise decisions in guideline production [294]*). The project manager has prepared a document based on a DECIDE template and distributed it to all participants in advance.. Nine members of the panel attend. Two of the absent members have been very active in previous meetings. Besides clinicians from several professions and organisations, staff from the Directorate of Health and a user representative, the panel also have had access to legal advice. In addition, two persons (hereafter called DECIDE advisors) who have had experience with the DECIDE processes in the Directorate of Health join the session. After coffee the project manager sets out the agenda and what needs to be discussed that day "whether our research-based, experience-based and user-based knowledge permits us to conclude that [screening tool A] is a better alternative than [screening tool B], or the opposite. If not, there will be a third option, meaning that we recommend or propose the use of either [screening tool A] or [screening tool B] as suitable tools. Thus, there can be three outcomes, [screening tool A], [screening tool B] or both.". This diverges from what was said in the first meeting, where the intention seemed to be to go for only one tool. The panel briefly discusses whether choosing one tool is actually important for the patient, but next, the two DECIDE advisors present the DECIDE method. The panel is to rank criteria (values, advantages versus disadvantages*

of possible interventions, resource use, equality, acceptability and feasibility) first and then compare the two tools.

The project manager leads the following discussion. The group starts to discuss what the subcategories of the values criterion (easy to understand, easy to use correctly, quick to apply, suitable for various patient groups, easy to communicate/implement and need for decision support?) actually mean.

When comparing the two tools based on the criteria, the nature of the discussion reveals that some panel members have formed a clear opinion regarding the tool they prefer, and the tools are no longer discussed systematically with respect to the separate DECIDE criteria.

A and B are clearly in favour of [screening tool A], while C [one of the assisting project leaders] and D clearly favour [screening tool B]. F attempts to prevent [screening tool A] from being rejected. C (assisting project manager) then says that she has to inform the panel that the hospital she works for as a clinician, already has decided that if the guideline ends up recommending [tool A], then this hospital will not follow the guideline's recommendation. The debate results in a quite strong attack from one of the DECIDE advisors on panel member assisting project leader C: "That's YOUR opinion". The DECIDE advisor then takes a much more active – and critical – role than previously. It almost seems as though she is trying to prevent assisting project leader C from gaining too much influence. The panel continues discussing and weighing the criteria using evidence mixed with personal experience on in what contexts and how the tools are used.

"I can say what I had planned to say a while ago. What I'm feeling intuitively is that [screening tool B] has been widely tested in hospitals, but when it comes to home nursing

etc. there are fewer data, because they have tested [screening tool C, another assessment tool]. As a result, we have a poor basis for concluding what will be best in nursing homes and home nursing, and I think it's difficult to require them to switch to [screening tool B]. After all, we don't know." Panel members discuss that in their experience some organisations will not be happy to implement other tools.

At 1500h the meeting ends, but there is a sort of weary and undefined mood in the room. Clear disagreement between the two key figures remains regarding the choice of instrument and whether having two instruments is acceptable. A and B are suddenly willing to retract their view that [screening tool A] is the best alternative, since they can see that the hospitals will not agree to use it. Some mumble about why they cannot create their own instrument and combine the question of illness with the questions in [screening tool A].

Chapter 10: Discussion and conclusions

This chapter contains updated versions of sections from papers co-authored with my supervisors and published elsewhere [16,90–92,296]*. In all cases, the main intellectual ideas were my own though my supervisors contributed to refinement and expression of those ideas.*

10.1 Introduction

In this chapter I revisit and reflect on the key themes in this thesis and how they address the main research question which involved examining tensions between guideline and mindlines considering Kuhn's ideas on anomalies within paradigms. I address the methodology and limitations of my research. Finally, I consider the implications of my research and provide recommendations and suggestions for further research and teaching EBM.

10.2 Summary of key findings

In this thesis I used several linked approaches to answer the main research question - How does the philosophy of knowledge help us to explore the tensions between guidelines and mindlines? In Chapter 2 I developed a Kuhn inspired perspective of guidelines as a central tool of EBM. In this view EBM is regarded as a paradigm consisting of beliefs, values and techniques that revolve around clinical decision support for

individual patients. The aim of this thesis was to find anomalies of that paradigm by questioning how guidelines could be re-tooled.

Kuhn described how 'normal science' is practiced in a mature paradigm. This means that researchers apply the assumptions, values and techniques of that paradigm to solve the typical scientific questions and problems that emerge in a scientific field. He calls these *puzzles.* Anomalies are findings that "*cannot, despite repeated effort, be aligned with professional expectation*" and phenomena *"for which his paradigm had not readied the investigator*"[14]. Awareness of anomalies is "*a recognition that nature has somehow violated the paradigm-induced expectations that govern normal science*" [14].

If EBM is considered to have matured from a new paradigm into 'normal science' over the past 25 years, anomalies would be the puzzles that cannot be solved by the beliefs (fundamental assumptions) and the central tools (guidelines) that the EBM paradigm currently provides for decision making support. In Chapter 4 I argued how the mindlines violate the expectations of the EBM paradigm regarding the nature of knowledge. Instead of universal, objective, proven, ready to be translated *facts-out-there* [297], EBM is challenged with a view on knowledge as subjective, tacit, fluid, dynamic and shared. Mindlines also raise questions on other epistemological topics, including reasoning and the nature of reality. Empirical research in Chapter 8 indicated that knowledge in informal communities of clinicians appears to better fit the concept of mindlines than how guideline panels handle knowledge as objective facts as addressed in Chapter 9.

In Chapter 5 I questioned whether the emergence of EBM movement marked a shift from the traditional 'priestly authority' of doctors to a professional expectation of a more modern, knowledge-based – and hence rational and scientific – authority. Using Latour's views on modernism I argued that recent developments showed that a changing EBM movement appears to acknowledge that this aim is ultimately unachievable. I considered how more reflexivity within the EBM community, a more sophisticated understanding of the translational process and an increasing interest in less purified kinds of knowledge seems to suggest a move towards a middle kingdom. Here reality is explained with reference to hybrids or the so-called quasi-objects. From this perspective, guidelines, in contrast to mindlines, are still limited to provide a holistic understanding of a real patient. EBM as a paradigm does stress the need for integration of external evidence and need to judge whether an evidence or recommendation applies to a specific patient (see Chapter 2) but appears to have prepared clinicians insufficiently for the paradox of a modernist agenda with clinical decision making for an individual patient.

In Chapter 6 I considered the anomalies perceived by the guideline developing community regarding reasoning and integrating different kinds of knowledge. Inspired by Ian Hacking's views on probability [28], I argued that EBM guideline development revolves around the frequency type evasion requiring explicit, statistical knowledge. As such, guidelines as central tools of EBM, offer clinicians few solutions to overcome the problem of induction to make inferences for a single patient in a specific situation. Other evasions however could be considered, including Bayesian reasoning, mechanistic reasoning, and non-analytical reasoning such as the use of gut feelings. The input for

these kinds of reasoning is a much wider range of knowledge as better encapsulated in the concept of mindlines. The finding of the guideline community that making good recommendations requires appraising and including different kinds of knowledge in guidelines, violates the expectation of the EBM paradigm of what inductive inferences could be considered valid.

In Chapter 7 I argued that EBM's struggle with providing adequate solutions for dealing with bias could originate from assumptions underlying the ideal limit theorem as a hidden theory of truth. Inspired by Isabelle Stenger's concept of *inter-esse*, I suggested how an alternative perspective on bias challenges EBM's assumptions. This view adds a productive dimension to bias as a drive to a certain line of questioning that cannot be eliminated. The duality that Stengers' view invokes draws attention to and challenges the assumptions underlying the ideal limit theory of truth in several ways. Most importantly, it casts doubt on the ideal limit theory as it applies to the single case scenario of the clinical encounter, the cornerstone of EBM. It indicates how EBM has not readied guideline makers sufficiently to understand human will in decision making other than in materialist terms. I theorised that in the complexity of clinical practice cognitive biases are contested and EBM could usefully incorporate a more pluralist understanding of truth instead of striving solely for the ideal limit concept.

Based on empirical findings presented in Chapter 9 I suggested how guideline development production can be understood as a practice performing a mixture of two contrasting epistemologies trying to reduce epistemic uncertainty. One a more purist position focusing on frequency type reasoning and the ideal limit theorem, and another

more pluralist position, applying multiple kinds of knowledge, types of reasoning and concepts of truth. I argue that these anomalies could have emerged from latent efforts to deal with anomalies of the EBM paradigm as discussed in this thesis. The inherent incommensurability between these two could serve as a further explanation of the tensions between formal guidelines and informal clinical mindlines.

Before I turn to potential ramifications of these anomalies, I first want to address the overall approach and methodologies used.

10.3 Reflections and limitations of methodology

In this thesis I took an interpretive approach to explore and reflect on the tensions between mindlines and guidelines from an ontological and epistemological perspective. I combined qualitative systematic review with empirical techniques including systematic reviewing, classic and digital ethnography, as well as undertaking analysis of concepts and theories from various philosophical perspectives.

When I started, I had no off-the-shelf frameworks to adopt. It appeared to me that EBM scholars and philosophers of science were working in different realms, unaware of each other's work. As a result, I found the interpretivist approach I chose helpful as it provided flexible methods to investigate the complexities of the creation of knowledge holistically from the in- and outside, without the burden of applying strict categorisations. In addition, I believe that participating as a GP in guideline development and experiencing

the issues first-hand helped my understanding. The risk that this could unduly have shaped my thesis, was counteracted by theoretical and methodological rigour. Part of the data was independently collected by two other observers, in one occasion on the same guideline panel. Analysis of data and deliberations of philosophical topics were always discussed with co-authors. I usually sought feedback on outcomes and conclusions from stakeholders prior to publication, for instance as part of the ethical protocol (digital ethnography) or institutional requirements of guideline making institutions. Drafts of the work reported in this thesis have also been presented at conferences (including e.g. Cochrane and Guideline International Network) for scrutiny. Furthermore, my ideas were shaped, and conclusions scrutinised through collaborations with guideline makers and scholars, most notably with the KNOWIT group at Oslo University, the CauseHealth initiative and the GIN AID working group.

As a critique of the EBM paradigm, my thesis had a risk of building a 'straw man' out of EBM due to the difficulty of delineating it as a concept. This might have partly originated from the shift in its definition from de-emphasing intuition, unsystematic clinical experience, and pathophysiologic rationale [20] towards a more inclusive focus on integration of external evidence with clinical expertise and patient preferences [3]. It might be caused by the experience with EBM in everyday practice, where it is often equated with the use of clinical guidelines [18] as discussed in Chapter 2. This could lead to a worry that EBM is critiqued as a limited practice with a myopic focus on statistical evidence, divorced from patient preferences and clinical expertise. Conversely, others have argued that EBM should be primarily seen as a research literacy movement,

empowering clinicians. They warn against an overgeneralised definition of EBM that implicates too many healthcare domains, giving rise to unfocused judgements [298].

In this light it is important to note that the understanding of EBM as critiqued in this thesis was framed according to its philosophical scope as laid out in Chapter 2. I have tried build a fair interpretation of the EBM paradigm by listening and collaborating with guideline developers who face the challenges of EBM in their work as seen in Chapter 6. If EBM is regarded as a primarily educational movement, the philosophy of science perspective led me to a focus on the contents of EBM, questioning the underlying assumptions of its teachings. Moreover, although there seems to be broad consensus that EBM grew out of clinical epidemiology [45,298,299], I have attempted to avoid judging EBM solely on how it handles external evidence from this perspective. Instead, I have emphasised its insistence on the integration of external evidence, patient preferences and clinical expertise by examining inductive inference and validity in making decisions for *individual* patients as discussed in Chapter 6 and 8.

However, I have to acknowledge that my work did not empirically analyse the guidelines-mindlines tension at the level of the clinical encounter. This has been explored by Gabbay and Le May's landmark work [10,15,300] and subsequently by others [301,302]. This thesis addressed the production of guidelines (in their role to support decision making for individual patients) rather than their actual use in the single case scenario and complements previous work by others in this way. Secondly, my study was not designed to provide detail on how these tensions are transmitted or 'intermediated' [155] from guideline production all the way down to supporting the

actual decision making of a single patient and a clinician. My focus on guideline development should not be interpreted as implying that the tensions between guidelines and mindlines are only significant at the higher levels of inception and development. And thirdly, from a philosophical perspective, I feel my theoretical work has sometimes only been able to scratch the surface of philosophical concepts as the primary intended audience for this work was clinicians and guideline makers, not academic philosophers. Whilst I believe this thesis has made some contributions to the philosophy of knowledge in healthcare, I recognise that there has been some trade-off between this and engaging clinicians and the guideline-making community.

Akin to my overall approach, each of the three primary research projects as reported in Chapters 4, 8 and 9 had particular strengths and limitations. A strength of the literature review In Chapter 4 as published [16] may have been its tight focus on the word 'mindlines' and Gabbay and Le May's original 2004 paper in the literature to address the question of how the word and their work have been used, and the meta-narrative approach which allowed us to consider different philosophical assumptions behind different uses of the word. A limitation is that the concept of mindlines may have been discussed more extensively in grey literature, institutional reports and other forums, which would have been detected using our search.

The online study on mindlines [92] in Chapter 8 had several limitations when considering the principles of digital ethnography as suggested by Pink et al. [83]. Regarding multiplicity, each network represented a different way in which clinicians could engage with the digital communities: an email-based network, a professional network and a

commercial network. As to non-digital-centric-ness, I did not examine how these virtual communities fit into the non-digital lives of their members. Other research would be necessary to do this. About openness, I experienced an informative interaction especially with the administrators and the members on the TGG group early on in the research. I was known to the community as a researcher in EBHC, but this appears unlikely to have led to any significant distortions in the content of the discussion as I did not participate actively in the discussions held. The contributions of the administrators and members however helped to shape the study to its final form. However, the HAWeb and Eyr communities were less responsive, possibly due to their lower activity. With regard to reflexivity, I acknowledge the importance of being critical of how the outcomes of this research were produced. Limitations of this study were the narrow scope of the topic and the limited number of virtual networks. Also, the defined timeframe may not have reflected how the activity (number of posts and people) on the networks changed over time. This makes it impossible to state confidently that the findings reflect online discussions generally rather than discussions on a particular topic. As such the findings may not be generalisable to digital mindlines more widely. A broader set of topics or other virtual networks may have revealed additional characteristics of knowledge on virtual networks. Concerning the principle of unorthodoxy, my aim was to continue the dialogue with the research participants, but I have not been able to gain more traction on this topic after publication of the study.

A strength of the ethnography of guideline panels in Chapter 9 would be its broad inclusion of kinds of topics, guideline panels in several countries and the use of participatory as well as observational techniques to gain a deep understanding of

guideline production. It may however be possible that certain guideline development setups and processes have not been witnessed as guideline making comes in many forms and shapes. However, I would argue that this would more likely further support the main conclusion that guideline making appears to be pluralistic than oppose it.

Bearing these cautions and limitations of the overall research approach and primary studies in mind I now turn to discussing the wider implications of my research.

10.4 Implications and further research

The scope of this thesis was confined to understanding how EBM reaches the limits of its paradigm in producing meaningful and valid guidelines. In this section I turn to what may lie beyond guideline development by professional institutions based on the anomalies of the EBM paradigm set out in this thesis.

It is important to note that Kuhn suggests that an anomaly per se is not sufficient to evoke a paradigm shift. Anomalies are part and parcel of a paradigm and sometimes they do get solved within the paradigm. Furthermore, contemplating all anomalies can be inefficient as normal science remains important. He writes:

> "*It follows that if an anomaly is to evoke crisis, it must usually be more than just an anomaly. There are always difficulties somewhere in the paradigm-nature fit;*

> *most of them are set right sooner or later, often by processes that could not have been foreseen. The scientist who pauses to examine every anomaly he notes will seldom get significant work done. We therefore have to ask what it is that makes an anomaly seem worth concerted scrutiny, and to that question there is probably no fully general answer.* [14]

In the case of EBM, the question is whether the anomalies suggested in this thesis are solvable within the paradigm. As described in Chapter 5, EBM as a movement is changing and trying to find solutions for the puzzles it faces. New developments can be helpful, but also pose new challenges. For instance, some of the changes in external evidence may be helpful, such as more inclusive trials, more registration and better quality control of trials, while others, such as testing similar drugs instead of finding new treatments, complicated protocols and rigid inclusion criteria [303] are not. Kuhn notes that:

> *"Progress in science is not a simple line leading to* the *truth. It is more progress away* from less *adequate conceptions of, and interactions with, the world"* [14]

Even when the anomalies are not solvable, it takes awareness of the scientific field to devote significant time and energy to it. But Kuhn proposes that once a set of anomalies is recognised to be more than one of the puzzles in normal science (and fundamentally question the explicit assumptions of the paradigm) a shift begins to form towards 'extraordinary science'.

> *"What is extraordinary research like? How is anomaly made lawlike? How do scientists proceed when aware only that something has gone fundamentally wrong at a level with which their training has not equipped them to deal? Those questions need far more investigation, and it ought not all be historical. What follows will necessarily be more tentative and less complete than what has gone before."* [14]

Here I want to consider what extraordinary science could mean in the context of clinical decision making. Below I suggest alternative assumptions to explore and tools to develop, bearing in mind that some of these could be acceptable and more easily fitted in the EBM paradigm, while others will be "more tentative and less complete".

From knowledge translation to mindlines[†]

The concept of mindlines has already inspired research with a new approach to knowledge translation and mobilisation [304,305]. Although the term 'mindlines' stuck as an antonym of 'guidelines', Gabbay and Le May who coined the term are the first to admit they are not very content with the term 'lines' [15]. In their book [15], they write that mindlines should not be seen as a measurable, closed system but *"[…] more like diffuse, often bending sets of influences which vary on different days in their impact depending on what else is going on and which sometimes go in different directions*". Equivalent to the inherent duality of knowledge artifacts (as holders or as catalysts of

[†] Parts of this section were previously published in [92]

knowledge), the authors have recently presented mindlines as processes, not entities per se [34,300].

The findings as presented in Chapters 4 and 8 would be consistent with seeing mindlines as an intermediate for explicit and tacit knowing shared among social groups and reinforcing norms of good practice in a fluid, dynamic and constantly evolving way. As such, mindlines can act as a pragmatic container concept, sufficiently vague to be able to dynamically hold many theories of knowledge. Instead of defined in terms of an entity (fact-out-there) or a property of an object (a single brain or computer) knowledge as in mindlines could be seen as a property of an interaction (like force and energy in physics): not directly observable but guiding and manifesting itself in real world decision making.

This suggests a move away from the conception of decision making in EBM. Instead of clear, predetermined preferences ready at hand, patients and clinicians would have implicit and unspecified preferences, that only become actual and clear in the process of making the decision. It challenges the use of decision aids and classic probability theory, which would require distinct and independent options [255,306]. This seems to be in line with Annemarie Mol's ideas on a logic of care [31]. As mindlines consider knowledge to be individually held and shared at the same time, and could play out in every context, they also question the possibility and necessity of dis-interest of researchers and guideline makers. It should be accepted that different panels will make different recommendations based on the same evidence, even in similar contexts [307]. Further research could develop new models and examine decision making in light of tensions between guidelines and mindlines on the level of the clinical encounter.

From guidelines to a new genre of tools

The findings in Chapter 9 suggest that, contrary to what is generally assumed, the guideline production process is not – and indeed cannot be – restricted to purely frequentist reasoning and objective judgements. To the extent that this is true (and that may vary considerably depending on the topic), the question arises of how to systematize and benchmark the inclusion of other forms of knowledge and reasoning (such as panel members' accumulated caseload of personal clinical experience and their intuitive or emotionally-driven hunches, collective deliberation, mechanistic reasoning and so on). Whilst caution is needed against any attempt to 'over-script' this complex and emergent process, a more pluralist stance, as promoted by many [240,285,308], is required. Guidelines could benefit, for example, from more insights into molecular and physiological mechanisms [259,309], qualitative studies to provide complementary perspectives [310], the influence of local contexts [311,312] and a risk assessment of recommendations being wrong. Ultimately, guideline production should be a reflexive practice, continuously critical and aware of its philosophical assumptions.

In Kuhnian terms, some re-tooling already took place within in the EBM paradigm, as knowledge tools to facilitate this have been developed. The "just in time" and "just enough" tools such as medication choice cards and option grid decision aids [313–315] but also NHG guidelines, are characterised by being short, flexible and accessible, address the reason for the clinical encounter (not a textbook condition) and they compare alternative options [313]. However, for new forms of guidelines to interact in a meaningful way with the mindlines that exist as informal knowledge in the community of clinicians (see Chapter 8), several further challenges would need to be dealt with to

support evidence-based decision making in the clinical encounter. In a paper produced during this thesis, my co-authors and I argued that not only should guidelines include other kinds of knowledge and reasoning, they also have to work with imperfect knowledge, be able to deal with unique single case scenarios, and deal with multiple options as well as multiple actions [313]. If and how this can be achieved exactly remains a topic for further research.

Virtual social networks linking guidelines to mindlines[‡]

To allow this new genre of tools to be produced fast and frugally and in collaborations with patients, many creative software formats have been developed, which contrast with guideline panels that typically take years to arrive at definitive recommendations [313]. However, their processes may not necessarily link them to mindlines of communities of clinicians or patients. The far less structured interactions on the virtual social networks in the research as discussed in chapter 8 represent a broad understanding of knowledge associated with many important knowledge theories that may be of use for guideline development. The findings show that not all networks will provide deep, rich knowledge. But the idea of analysing selected virtual social networks as part of guideline update processes could help to better frame and synchronize recommendations with the mindlines of clinicians. This could be considered 'extraordinary research' in the Kuhnian sense and immediately pose some basic and fundamental questions: How would we define a virtual social network? What does a good network look like? How should we study these networks?

[‡] Parts have previously been published in [92]

Some principles of internet research could help to develop an appropriate methodology to do this. As Christine Hine wrote "Internet research has arguably been a valuable reflexive opportunity for the traditional disciplines that have fed its development"[316]. Guideline developers could consider experimenting and set up collaborations with these networks. Further research would be required to examine whether using virtual social networks in this way in guideline production processes is feasible and produces meaningful guidance.

Towards a dialogue on purification, inference and epistemological uncertainty[§]

At the same time, new knowledge tools and virtual networks remain purifications of reality, unable to fully substitute for judgment, experience, sensitivity, and humanity [313] of a real face-to-face encounter in clinical practice. A dialogue within the EBM movement is needed about the role, the paradox, benefits and harms of the modernist practice of purification. With Wood I would argue for more interdisciplinary practices: "innovation is neither natural or inevitable, but constantly negotiated and aligned – a path forged within an assemblage of scientific and organisational and behavioural factors" [187]. Rather than a sequence of research-driven abstract decisions, inference making beyond EBM should be conceptualised as a situated practice, always starting from the real problems of individual real patients (and real organisations and systems) in their real contexts.

[§] Parts were previously published in [89]

This dialogue should already start in medical school teaching students about epistemological uncertainty which lies at the heart of what it is to practice excellent medicine [317]. Much of the training we offer students in EBM however can be thought of as a systematic approach to undertaking Latour's purification (as in reducing the patient's presenting problem to a series of PICO questions). Critics argue that this training encourages them to over-value knowledge that can be made explicit and quantified, ready for frequentist induction. The findings in Chapter 9 of this thesis have shown that other epistemologies are at play.

Instead of a limited practice of 'evidence-basing', a paradigm beyond EBM could concern itself with 'inference in healthcare' as the many ways to explicate and mitigate epistemic uncertainty [102]. This would include a broader understanding of causality, including dispositions [204,257,318], final cause and non-local causation [319]. In the process of building capability to cope with clinical complexity ('the ability to adapt to change, generate new knowledge, and continuously improve performance') [320], students should at least be able to debate and consider some of the broader philosophical assumptions underpinning inductive inference in healthcare. Furthermore, EBM teaching could give heed to Hasok Chang's work [183] as noted in Chapter 4 and reassess curricula and theory from before EBM to learn from the approaches taken by previous generations of educators.

From frequency reasoning to situated inference

As discussed in Chapter 6, RCTs and guidelines largely based on a hierarchy of evidence [3,19,48] are only able to provide inferences for groups of patients directly, not in single

case scenarios (e.g. a clinician seeing one patient), when there is no frequency of events. For this reason, it seems inherently problematic to adhere to rule-based reasoning from the general (group derived recommendations) to the individual. Alternative, more complex ways of reasoning [306,321–323], may very well become the dominant tools of a paradigm beyond EBM, replacing both RTCs and guidelines. Further research would be required to understand their philosophical assumptions, and how they can (or cannot) support decision making in a meaningful way. Yet, besides the challenge of acceptability of the hidden and complex reasoning they convey, these contenders might remain limited to the codified, purified data they can process [313].

To overcome this problem further research could consider whether direct individual-to-individual reasoning (casuistry instead of rule-based reasoning) [203] is able to derive meaningful inferences in the individual scenario. Case-based reasoning seems not to have been a prominent way of reasoning since the emergence of EBM. Case reports have suffered from a low status on the hierarchy of evidence. However, case report publication is yet again on the rise [324]. It has been suggested this is to describe new diseases, to detect side effects, to study mechanisms, for medical education or recognition of rare events [325]. But there may be other processes at play and further research theorisation is necessary to analyse their role in inference. Again, as this topic used to be researched all the time, the first task might be to do a historical review of what has been published in the past.

From hiding behind bias to critically appraising truth and pluralism[**]

As suggested in Chapter 7, in order to deal with the anomaly of bias, EBM may need to shift from a narrow focus on bias based on the ideal limit concept of truth towards more an extended understanding of many theories of truth. This kind of thinking about other truth concepts is still preliminary and more work is needed to develop a more detailed comprehension of bias and truth in clinical decision making. Further research would help to provide an underlying theory to qualify the perceived bias, to understand how the ideal limit concept works in reality and to further study how the different truth concepts interfere with each other [274].

It may be possible to solve the challenge of dealing with bias as a normal puzzle within the EBM paradigm. By making future clinicians more literate [298] about underlying theories of truth the ideal limit theorem could possibly in some extent be reconciled with human interests in the clinical encounter. However, the controversy of two contrasting epistemological practices in guideline development as a response to the many anomalies of the EBM paradigm as noted in Chapter 9, may indicate that a more radical shift is needed. This shift may have to go beyond critiquing EBM for its positivist position [52].

Guideline development as an unavoidably pluralist practice may be a necessary but at the same time unsatisfactory position for many, and not just for classic EBM proponents. A subliminal longing for a more connected, less disjointed epistemology (under the

[**] Parts have previously been published in [91]

assumption of a single reality) seems to persist. As discussed in Chapter 2, the central goal of EBM was defined as *integrating* the best evidence with the clinical expertise and patient preferences [3]. As already quoted in in Chapter 4, Malterud noticed that the EBM movement does not limit the best evidence to RCTs and meta-analysis,

> *"Yet, the foundation for integrating the available sources of knowledge remains unclear. We still do not know whether convincing information leads to optimal decision making* [89]."

Similarly, Jonas argued that *"we need to broaden and deepen our understanding of what counts as "evidence" and which types of evidence are best used to inform differing aspects of clinical decision making* [90]". But a foundation for integration of evidence or a basis to believe that specific types of evidence link better to specific issues than others is not what pluralism has on offer [326]. Further research would be needed to critique pluralism. This could potentially result in alternative philosophical positions to guide inductive inference in clinical practice.

10.5 Final remarks

This thesis originated from an interest to gain a better understanding of clinical guideline development as well as to find out what may lie beyond EBM. My empirical and theoretical work was aimed at providing new insights in both areas using an epistemological perspective, with a focus on knowledge, reasoning, truth and purification. I examined tensions between mindlines and guidelines as anomalies of a

mature Kuhnian paradigm. Most notably my research suggested a tension between two linked yet contrasting epistemologies within the guideline community: a purist understanding of EBM, adhering to frequency type reasoning, the ideal limit theorem and strict explicit purification, and a more pluralist understanding of EBM embracing multiple kinds of knowledge, reasoning and truth.

My research has taken me to a broad exploration of some of the major issues in the philosophy of science and altered my views on guideline making and EBM. As a practising GP, it had an impact on my decision making as it made me question my own philosophical assumptions and framings. As a result, I turned from trying to find technical solutions to make better recommendations to being more reflexive and critical about the fundamentals of clinical knowledge creation.

With Tonelli et al, I believe that the philosophy of science is only at the start of emerging as a academic field in healthcare [102]. In depth considerations of the concepts that underlie inductive inference will help guide EBM's transformation into something new. It seems inescapable that the paradigm originally proposed by Guyatt et al. [20] fades when knowledge creation becomes understood in a more nuanced way. Finding out what lies beyond EBM calls for work and imagination from interested scholars, clinicians, guideline developers and patients.

Literature

1 Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? *BMJ Br Med J* 2014;**348**. doi:10.1136/bmj.g3725

2 Zuiderent-Jerak T, Forland F, Macbeth F. Guidelines should reflect all knowledge, not just clinical trials. *BMJ* 2012;**345**:e6702. doi:10.1136/bmj.e6702

3 Sackett DL, Rosenberg WMC, Gray JAM, *et al.* Evidence based medicine: what it is and what it isn't. *BMJ* 1996;**312**:71–2. doi:10.1136/bmj.312.7023.71

4 Tonelli MR. The philosophical limits of evidence-based medicine. *Acad Med* 1998;**73**:1234–40.

5 Cohen AM, Stavri PZ, Hersh WR. A categorization and analysis of the criticisms of Evidence-Based Medicine. *Int J Med Inform* 2004;**73**:35–43. doi:10.1016/j.ijmedinf.2003.11.002

6 Raab M, Gigerenzer G. The power of simplicity: a fast-and-frugal heuristics approach to performance science. *Front Psychol* 2015;**6**:1–6. doi:10.3389/fpsyg.2015.01672

7 Greenhalgh T, Snow R, Ryan S, *et al.* Six 'biases' against patients and carers in evidence-based medicine. *BMC Med* 2015;**13**:200. doi:10.1186/s12916-015-0437-x

8 Sniderman AD, Furberg CD. Why guideline-making requires reform. *JAMA* 2009;**301**:429–31. doi:10.1001/jama.2009.15

9 Beauchamp S, Drapeau M, Dionne C. The development of practice guidelines in the social and human sciences. *Can Psychol Can* 2015;**56**:357–67. doi:10.1037/a0038811

10 Gabbay J, Le May A. Evidence based guidelines or collectively constructed 'mindlines?' Ethnographic study of knowledge management in primary care. *BMJ Br Med J* 2004;**329**:1013. doi:10.1136/bmj.329.7473.1013

11 Currie G, Burgess N, White L, *et al.* A qualitative study of the knowledge-brokering role of middle-level managers in service innovation: managing the translation gap in patient safety for older persons' care. *Heal Serv Deliv Res* 2014;**2**:1–118. doi:10.3310/hsdr02320

12 Greenhalgh T, Robert G, Macfarlane F, *et al.* Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Soc Sci Med* 2005;**61**:417–30. doi:10.1016/j.socscimed.2004.12.001

13 Wong G, Greenhalgh T, Westhorp G, *et al.* RAMESES publication standards: meta-narrative reviews. *BMC Med* 2013;**11**:20. doi:10.1186/1741-7015-11-20

14 Kuhn TS. *The Structure of Scientific Revolutions*. The University of Chicago Press 1970.

15 Gabbay J, Le May A. *Practice-Based Evidence for Healthcare: Clinical Mindlines*. Abingdon: : Routledge 2011.

16 Wieringa S, Greenhalgh T. 10 years of mindlines: a systematic review and commentary. *Implement Sci* 2015;**10**:45. doi:10.1186/s13012-015-0229-x

17 Glasziou P, Mar C Del, Salisbury J. *Evidence based medicine workbook*. BMJ Publishing Group 2003. http://eprints.kmu.ac.ir/2228/1/Evidence-Based Medicine Workbook - 1st Edition.pdf (accessed 19 Nov 2013).

18 Timmermans S, Berg M. *The Gold Standard*. Philadelphia: : Temple University Press 2003.

19 Howick JH. *The Philosophy of Evidence-based Medicine*. BMJ Books. Wiley-Blackwell 2011.
20 Evidence-Based Medicine Working Group. Evidence-Based Medicine A New Approach to Teaching the Practice of Medicine. *Jama* 1992;**268**:2420–5.
21 Latour B. *We Have Never Been Modern*. Cambridge: : Harvard University Press 1993. doi:10.1016/0956-5221(96)88504-6
22 Loder E, Groves T. The BMJ requires data sharing on request for all trials. *BMJ* 2015;**350**:h2373. doi:10.1136/bmj.h2373
23 AllTrials. All Trials Declaration. 2013.
24 Edwards A, Elwyn G. Shared decision-making in health care : Achieving evidence-based patient choice. 2009;:3–10.
25 Ekman I, Swedberg K, Taft C, *et al.* Person-centered care--ready for prime time. *Eur J Cardiovasc Nurs* 2011;**10**:248–51. doi:10.1016/j.ejcnurse.2011.06.008
26 Elwyn G, Quinlan C, Mulley A, *et al.* Trustworthy guidelines - excellent; customized care tools - even better. *BMC Med* 2015;**13**:199. doi:10.1186/s12916-015-0436-y
27 Hume D. *A Treatise of Human Nature*. Google eBo. London: : Thomas and Joseph Allman 1817. http://books.google.com/books?id=-L4IAAAAQAAJ&pgis=1
28 Hacking I. *An Introduction to Probability and Inductive Logic*. Cambridge university press 2001.
29 Peirce C. *Philosophical Writings*. Dover Publications Inc. 1986.
30 Popper KR. *Objective Knowledge: An Evolutionary Approach*. Clarendon Press 1979.
31 Mol A. *The Logic of Care: Health and the Problem of Patient Choice*. Routledge 2008. http://books.google.com/books?id=w2IV2jwKFa4C&pgis=1 (accessed 3 Oct 2013).
32 Stengers I. *Power and Invention: Situating Science*. Min nea polis: : University of Minnesota Press 1997.
33 Cabitza F, Colombo G, Simone C. Leveraging underspecification in knowledge artifacts to foster collaborative activities in professional communities. *Int J Hum Comput Stud* 2013;**71**:24–45. doi:10.1016/j.ijhcs.2012.02.005
34 Cabitza F. At the Boundary of Communities and Roles: Boundary Objects and Knowledge Artifacts as Resources for IS Design. In: *Lecture Notes in Information Systems and Organisation*. 2015. 149–60. doi:10.1007/978-3-319-09450-2_13
35 Steup M. Epistemology. Stanford Encycl. Philos. Summer 2018 Ed. Edward N. Zalta. 2018.https://plato.stanford.edu/archives/sum2018/entries/epistemology/
36 Bird A. Thomas Kuhn. Stanford Encycl. Philos. (Fall 2013 Ed. 2013.https://plato.stanford.edu/archives/fall2013/entries/thomas-kuhn
37 Sismondo S. *An Introduction to Science and Technology Studies*. Chicester: : John Wiley & Sons 2011.
38 Kuhn TS. Objectivity, Value Judgment, and Theory Choice. In: *The Essential Tension*. 1977. 390.http://shamiller.net/phi031/readings.d/kuhn.pdf
39 Massimi M. 'Working in a new world': Kuhn, constructivism, and mind-dependence. *Stud Hist Philos Sci Part A* 2015;**50**:83–9. doi:10.1016/j.shpsa.2014.09.011
40 Fleck L. *Genesis and Development of a Scientific Fact*. Chicago, London: : The

University of Chicago Press 1979.
41 Sak J, Pawlikowski J. Medicine and thought-styles: On the 50th anniversary of the death of Ludwik Fleck (1896-1961). *Isr Med Assoc J* 2012;**14**:214–8.
42 Carifio J, Perla RJ. Not Just a 'Fleck' on the Epistemic Landscape: A Reappraisal of Ludwik Fleck's Views of the Nature of Scientific Progress and Change in Relation to Contemporary Educational and Social Issues. *Res Sci Educ* 2013;**43**:2349–66. doi:10.1007/s11165-013-9361-2
43 Weiss K. Ludwik Fleck and the Art-of-Fact. *Evol Anthropol* 2003;**12**:168–72. doi:10.1002/evan.10118
44 Foucault M. *Power/Knowledge*. New York: : Pantheon Books 1980.
45 Bolt TC. *A Doctor's Order The Dutch Case of Evidence-Based Medicine (1970-2015)*. 2015.
46 Couto M D JS, Sa Couto J. Evidence-based medicine: a Kuhnian perspective of a transvestite non- theory. *J Eval Clin Pract* 1998;**4**:267–75. doi:10.1111/j.1365-2753.1998.tb00085.x
47 Sehon SR, Stanley DE. A philosophical analysis of the evidence-based medicine debate. *BMC Health Serv Res* 2003;**3**:14. doi:10.1186/1472-6963-3-14
48 Guyatt GH, Oxman AD, Vist GE, *et al.* GRADE : An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations. *Br Med J* 2008;**336**:924–6. doi:10.1136/bmj.39489.470347.AD
49 McMahon GT, Dluhy RG. Impact of a novel evidence-based medicine curriculum. *Med Educ* 2006;**40**:464–5. doi:10.1111/j.1365-2929.2006.02457.x
50 Toulmin S. Does the distinction between normal and revolutionary science hold water? *Crit Growth Knowl* 1970;:39–47. doi:10.1017/CBO9781139171434.005
51 Smith R, Rennie D, Chalmers TC, *et al.* Evidence-Based Medicine — An Oral History. 2014;**311**:1–3.
52 Djulbegovic B, Guyatt GH. Epistemologic Inquiries in Evidence-Based Medicine. *Cancer Control* 2009;**16**.
53 Bluhm R. Jeremy Howick: The philosophy of evidence-based medicine. *Theor Med Bioeth* 2011;**32**:423–7. doi:10.1007/s11017-011-9196-7
54 Meehl PE. *Clinical verus Statistical Prediction*. University of Minnesota 1996. doi:10.1902/jop.1999.70.6.700
55 Grove WM, Zald DH, Lebow BS, *et al.* Clinical versus mechanical prediction: A meta-analysis. *Psychol Assess* 2000;**12**:19–30. doi:10.1037/1040-3590.12.1.19
56 Bachmann LM, Kolb E, Koller MT, *et al.* Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: Systematic review. *Br Med J* 2003;**326**:417–9. doi:10.1136/bmj.326.7386.417
57 Solomon M. *Making Medical Knowledge*. Oxford University Press 2015.
58 Solomon M. Just a paradigm: Evidence-Based Medicine in epistemological context. *Eur J Philos Sci* 2011;**1**:451–66. doi:10.1007/s13194-011-0034-6
59 Rosner AL. Evidence-based medicine: revisiting the pyramid of priorities. *J Bodyw Mov Ther* 2012;**16**:42–9. doi:10.1016/j.jbmt.2011.05.003
60 Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs* 2003;**12**:77–84. doi:10.1046/j.1365-2702.2003.00662.x
61 Weisz G. From Clinical Counting to Evidence-Based Medicine. *Body Counts Med Quantif Hist Sociol Perspect Hist Sociol sur la Quantif m□dicale* 2005;:377–93.

62 Sanders S, Doust J, Glasziou P. A systematic review of studies comparing diagnostic clinical prediction rules with clinical judgment. *PLoS One* 2015;**10**:1–25. doi:10.1371/journal.pone.0128233
63 Walach H, Falkenberg T, Fønnebø V, *et al.* Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol* 2006;**6**:29. doi:10.1186/1471-2288-6-29
64 Henry SG. Recognizing tacit knowledge in medical epistemology. *Theor Med Bioeth* 2006;**27**:187–213. doi:10.1007/s11017-006-9005-x
65 Kelly MP, Heath I, Howick J, *et al.* The importance of values in evidence-based medicine. *BMC Med Ethics* 2015;**16**:69. doi:10.1186/s12910-015-0063-3
66 Loughlin M, Bluhm R, Buetow S, *et al.* Reason and value: Making reasoning fit for practice. *J Eval Clin Pract* 2012;**18**:929–37. doi:10.1111/j.1365-2753.2012.01896.x
67 Chalmers I, Bracken MB, Djulbegovic B, *et al.* How to increase value and reduce waste when research priorities are set. *Lancet* 2014;**383**:156–65. doi:10.1016/S0140-6736(13)62229-1
68 Gupta M. A critical appraisal of evidence-based medicine: Some ethical considerations. *J Eval Clin Pract* 2003;**9**:111–21. doi:10.1046/j.1365-2753.2003.00382.x
69 Gupta M. *Is Evidence-Based Psychiatry Ethical?* 2014.
70 Gigerenzer G, Brighton H. Homo Heuristicus: Why Biased Minds Make Better Inferences. *Top Cogn Sci* 2009;**1**:107–43. doi:10.1111/j.1756-8765.2008.01006.x
71 Montgomery K. *How Doctors Think. Clincal Judgement and the Practice of Medicine*. Oxford University Press 2006.
72 Engebretsen E, Vøllestad NK, Wahl AK, *et al.* Unpacking the process of interpretation in evidence-based decision making. *J Eval Clin Pract* 2015;**21**:529–31.http://blogs.lse.ac.uk/impactofsocialsciences/2015/05/08/interpretation-in-evidence-based-medicine/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+feedburner/LSEImpactBlog+(LSE+Impact+Blog)
73 Kahneman D, Tversky A. Choices , Values , and Frames. *Am Psychol* 1984;**39**:341–50.
74 Brożek JL, Akl EA, Alonso-Coello P, *et al.* Grading quality of evidence and strength of recommendations in clinical practice guidelines. *Allergy* 2009;**64**:669–77. doi:10.1111/j.1398-9995.2009.01973.x
75 Solution a P. Perspective Annals of Internal Medicine The Vexing Problem of Guidelines and Conflict of Interest : 2014;:738–42.
76 Mason J. *Qualitative Researching*. London: : SAGE Publications 2002.
77 Silverman D, editor. *Qualitative Research*. London: : SAGE Publications 2016.
78 Denzin, Norman K ; Lincoln YS. *The SAGE Handbook of Qualitative Research*. SAGE Publications 2017.
79 Hacking I. 'Language, Truth and Reason' 30 years later. *Stud Hist Philos Sci Part A* 2012;**43**:599–609. doi:10.1016/j.shpsa.2012.07.002
80 Wong G, Greenhalgh T, Westhorp G, *et al.* RAMESES publication standards: realist syntheses. *BMC Med* 2013;**11**:21. doi:10.1186/1741-7015-11-21
81 Madson M. Digital ethnography for intercultural professional communication : Some best practice principles. *Rhetor Prof Commun Glob* 2014;**5**:67–89.

82 Murthy D. An Examination of the Use of New Technologies for Social Research. *Sociology* 2008;**Vol. 42**:837–55. doi:10.1177/0038038508094565
83 Pink S, Horst H, Postill J, *et al.* Ethnography in a Digital World. In: *Digital Ethnography : Principles and Practices*. SAGE Publications 2016.
84 Smith J. From optimism to hubris. *BMJ Br Med J* 2004;**329**:0-h.
85 Francke AL, Smit MC, de Veer AJE, *et al.* Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review. *BMC Med Inform Decis Mak* 2008;**8**:38. doi:10.1186/1472-6947-8-38
86 Lugtenberg M, Zegers-van Schaick JM, Westert GP, *et al.* Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. *Implement Sci* 2009;**4**:54. doi:10.1186/1748-5908-4-54
87 Carlsen B, Glenton C, Pope C. Thou shalt versus thou shalt not : a meta-synthesis of GPs ' attitudes to clinical practice guidelines B Carlsen , C Glenton and C Pope. 2007;:971–8.
88 Orlikowski WJ, Iacono CS. Research Commentary : Desperately Seeking the " IT " in IT Research — A Call to Theorizing the IT Artifact. *Inf Syst Res* 2001;**12**:121–34.
89 Wieringa S, Engebretsen E, Heggen K, *et al.* Has evidence-based medicine ever been modern? A Latour-inspired understanding of a changing EBM. *J Eval Clin Pract* 2017;:1–7. doi:10.1111/jep.12752
90 Wieringa S, Dreesens D, Forland F, *et al.* Different knowledge, different styles of reasoning: a challenge for guideline development. *BMJ Evidence-Based Med* 2018;**0**:ebmed-2017-110844. doi:10.1136/bmjebm-2017-110844
91 Wieringa S, Engebretsen E, Heggen K, *et al.* Rethinking bias and truth in evidence-based health care. *J Eval Clin Pract* 2018;:1–9. doi:10.1111/jep.13010
92 Wieringa S, Engebretsen E, Heggen K, *et al.* How Knowledge Is Constructed and Exchanged in Virtual Communities of Physicians: Qualitative Study of Mindlines Online. *J Med Internet Res* 2018;**20**:e34. doi:10.2196/jmir.8325
93 Finegold JA, Shun-Shin MJ, Cole GD, *et al.* Distribution of lifespan gain from primary prevention intervention. *Open Hear* 2016;**3**:e000343. doi:10.1136/openhrt-2015-000343
94 Nissen SE, Stroes E, Dent-Acosta RE, *et al.* Efficacy and Tolerability of Evolocumab vs Ezetimibe in Patients With Muscle-Related Statin Intolerance: The GAUSS-3 Randomized Clinical Trial. *Jama* 2016;**315**:1580–90. doi:10.1001/jama.2016.3608
95 Collins R, Reith C, Emberson J, *et al.* Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet* 2016;**388**:2532–61. doi:10.1016/S0140-6736(16)31357-5
96 Lorgeril M De, Rabaeus M. Beyond confusion and controversy, can we evaluate the real efficacy and safety of cholesterol-lowering with statins? *J Controv Biomed Res* 2016;**1**:67. doi:10.15586/jcbmr.2015.11
97 Redberg RF, Katz MH. Statins for Primary Prevention. *JAMA Intern Med* 2017;**177**:21. doi:10.1001/jamainternmed.2016.7585
98 Buchanan EA, Ess CM, Markham AN. International Journal of Internet Research Ethics 3.1. 2010;**3**.
99 Markham A, Buchanan E. Ethical Decision-Making and Internet Research :

Version 2 . 0 Recommendations from the AoIR Ethics Working Committee. 2012;:1–19.
100 Sandelowski M, Barroso J. Classifying the findings in qualitative studies. *Qual Heal Res* 2003;**13**:905–23. doi:10.1177/1049732303253488
101 Hacking I. *Historical Ontology*. Cambridge: : Harvard University Press 2002.
102 Tonelli MR, Upshur REG. A Philosophical Approach to Addressing Uncertainty in Medical Education. *Acad Med* 2019;**94**:1. doi:10.1097/acm.0000000000002512
103 Djulbegovic B, Hozo I, Greenland S. *Uncertainty in Clinical Medicine*. Elsevier B.V. 2011. doi:10.1016/B978-0-444-51787-6.50011-8
104 Engebretsen E, Heggen K, Wieringa S, *et al.* Uncertainty and objectivity in clinical decision making: a clinical case in emergency medicine. *Med Health Care Philos* Published Online First: 2016. doi:10.1007/s11019-016-9714-5
105 Eddy DM. Variations in Physician Practice The Role of Uncertainty. *Health Aff* 1984;**3**:74–89. doi:10.4103/0970-0218.112447
106 Moreira T. Diversity in clinical guidelines: The role of repertoires of evaluation. *Soc Sci Med* 2005;**60**:1975–85. doi:10.1016/j.socscimed.2004.08.062
107 Calderón C, Rotaeche R, Etxebarria A, *et al.* Gaining insight into the Clinical Practice Guideline development processes: qualitative study in a workshop to implement the GRADE proposal in Spain. *BMC Health Serv Res* 2006;**6**:138. doi:10.1186/1472-6963-6-138
108 Atkins D, Briss P a, Eccles M, *et al.* Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res* 2005;**5**:25. doi:10.1186/1472-6963-5-25
109 Nairn S, Timmons S. Scientific uncertainty and the creation of resuscitation guidelines. *Soc Theory Heal* 2010;**8**:289–308. doi:10.1057/sth.2009.4
110 Knaapen L, Cazeneuve H, Cambrosio A, *et al.* Pragmatic evidence and textual arrangements: A case study of French clinical cancer guidelines. *Soc Sci Med* 2010;**71**:685–92. doi:10.1016/j.socscimed.2010.05.019
111 Palinkas LA, Horwitz SM, Green CA, *et al.* Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Adm Policy Ment Heal Ment Heal Serv Res* 2015;**42**:533–44. doi:10.1007/s10488-013-0528-y
112 Burgers JS, Grol R, Klazinga NS, *et al.* Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *Int J Qual Heal care* 2003;**15**:31–45.
113 Qaseem A, Forland F, Macbeth F, *et al.* Guidelines International Network: Toward International Standards for Clinical Practice Guidelines. *Ann Intern Med* 2012;**156**:525--U103. doi:10.1059/0003-4819-156-7-201204030-00009
114 Christiaens T, De Backer D, Burgers J, *et al.* Guidelines, evidence, and cultural factors. *Scand J Prim Health Care* 2004;**22**:141–5. doi:10.1080/02813430410006521
115 Glaser BG, Strauss AL. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. 1967. doi:10.2307/2575405
116 Mol A. *The Body Multiple: Ontology in Medical Practice*. Duke University Press 2002.
117 Winter R. Fictional-critical writing: An approach to case study research by practitioners. *Cambridge J Educ* 1986;**16**:175–82.

doi:10.1080/0305764860160303
118 Hesse-Biber S. Mixed methods research: The 'thing-ness' problem. *Qual Health Res* 2015;**25**:775–88. doi:10.1177/1049732315580558
119 Silversides A. Patient-safety reforms inhibited by systemic impediments. *CMAJ* 2008;**179**:1253–5. doi:10.1503/cmaj.081790
120 Moja L, Banzi R. Navigators for medicine: evolution of online point-of-care evidence-based services. *Int J Clin Pract* 2011;**65**:6–11. doi:10.1111/j.1742-1241.2010.02441.x
121 Gupta S, Bhattacharyya OK, Brouwers MC, *et al.* Canadian Thoracic Society: Presenting a new process for clinical practice guideline production. *Can Respir J* 2009;**16**:e62--68.
122 Gagliardi AR, Wright FC, Davis D, *et al.* Challenges in multidisciplinary cancer care among general surgeons in Canada. *BMC Med Inform Decis Mak* 2008;**8**:59. doi:10.1186/1472-6947-8-59
123 Kothari A, Rudman D, Dobbins M, *et al.* The use of tacit and explicit knowledge in public health: a qualitative study. *Implement Sci* 2012;**7**:20. doi:10.1186/1748-5908-7-20
124 Grant A, Sullivan F, Dowell J. An ethnographic exploration of influences on prescribing in general practice: why is there variation in prescribing practices? *Implement Sci* 2013;**8**:72. doi:10.1186/1748-5908-8-72
125 Elstad EA, Lutfey KE, Marceau LD, *et al.* What do physicians gain (and lose) with experience? Qualitative results from a cross-national study of diabetes. *Soc Sci Med* 2010;**70**:1728–36. doi:10.1016/j.socscimed.2010.02.014
126 Chandler CIR, Jones C, Boniface G, *et al.* Guidelines and mindlines: why do clinical staff over-diagnose malaria in Tanzania? A qualitative study. *Malar J* 2008;**7**:53. doi:10.1186/1475-2875-7-53
127 Li LC, Grimshaw JM, Nielsen C, *et al.* Evolution of Wenger's concept of community of practice. *Implement Sci* 2009;**4**:11. doi:10.1186/1748-5908-4-11
128 Argyris C, Schon D. *Organizational Learning II: Theory, Method, and Practice*. Reading, MA: : Addison-Wesley Publishing Company 1996.
129 Crites GE, McNamara MC, Akl EA, *et al.* Evidence in the learning organization. *Health Res Policy Syst* 2009;**7**:4. doi:10.1186/1478-4505-7-4
130 Ranmuthugala G, Plumb JJ, Cunningham FC, *et al.* How and why are communities of practice established in the healthcare sector? A systematic review of the literature. *BMC Health Serv Res* 2011;**11**:273. doi:10.1186/1472-6963-11-273
131 Soubhi H, Bayliss EA, Fortin M, *et al.* Learning and caring in communities of practice: using relationships and collective learning to improve primary care for patients with multimorbidity. *Ann Fam Med* 2010;**8**:170–7. doi:10.1370/afm.1056
132 Mascia D, Cicchetti A, Damiani G. 'Us and Them': a social network analysis of physicians' professional networks and their attitudes towards EBM. *BMC Health Serv Res* 2013;**13**:429. doi:10.1186/1472-6963-13-429
133 Barnett ML, Landon BE, O'Malley AJ, *et al.* Mapping Physician Networks with Self-Reported and Administrative Data. *Health Serv Res* 2011;**46**:1592–609. doi:10.1111/j.1475-6773.2011.01262.x
134 Barnett ML, Christakis NA, O'Malley J, *et al.* Physician patient-sharing networks

and the cost and intensity of care in US hospitals. *Med Care* 2012;**50**:152–60. doi:10.1097/MLR.0b013e31822dcef7
135 Keating NL, Ayanian JZ, Cleary PD, *et al.* Factors affecting influential discussions among physicians: a social network analysis of a primary care practice. *J Gen Intern Med* 2007;**22**:794–8. doi:10.1007/s11606-007-0190-8
136 Lomas J. The in-between world of knowledge brokering. *BMJ* 2007;**334**:129–32. doi:10.1136/bmj.39038.593380.AE
137 Oduro-Mensah E, Kwamie A, Antwi E, *et al.* Care decision making of frontline providers of maternal and newborn health services in the greater Accra region of Ghana. *PLoS One* 2013;**8**:e55610. doi:10.1371/journal.pone.0055610
138 Lucchiari C, Pravettoni G. Cognitive balanced model: A conceptual scheme of diagnostic decision making. J. Eval. Clin. Pract. 2012;**18**:82–8. doi:10.1111/j.1365-2753.2011.01771.x
139 Maskrey N, Underhill J, Hutchinson a., *et al.* Getting a better grip on research: A simple system that works. InnovAiT. 2009;**2**:739–49. doi:10.1093/innovait/inp106
140 Bate L, Hutchinson A, Underhill J, *et al.* How clinical decisions are made. *Br J Clin Pharmacol* 2012;**74**:614–20. doi:10.1111/j.1365-2125.2012.04366.x
141 Swayne J. The problem with science-the context and process of care: an excerpt from remodelling medicine. *Glob Adv Health Med* 2012;**1**:78–87. doi:10.7453/gahmj.2012.1.1.014
142 Contandriopoulos D, Lemire M, Denis J-L, *et al.* Knowledge exchange processes in organizations and policy arenas: a narrative systematic review of the literature. *Milbank Q* 2010;**88**:444–83. doi:10.1111/j.1468-0009.2010.00608.x
143 Glasziou P. Evidence based medicine: does it make a difference?: Make it evidence informed practice with a little wisdom. *BMJ Br Med J* 2005;**330**.
144 Rycroft-Malone J, Seers K, Chandler J, *et al.* The role of evidence, context, and facilitation in an implementation trial: implications for the development of the PARIHS framework. *Implement Sci* 2013;**8**:28. doi:10.1186/1748-5908-8-28
145 Jiwa M, Deas K, Ross J, *et al.* An inclusive approach to raising standards in general practice: working with a 'community of practice' in Western Australia. *BMC Med Res Methodol* 2009;**9**:13. doi:10.1186/1471-2288-9-13
146 Clark F, Park DJ, Burke JP. Dissemination: bringing translational research to completion. *Am J Occup Ther* 2013;**67**:185–93. doi:10.5014/ajot.2013.006148
147 Genuis SJ, Schwalfenberg GK, Hiltz MN, *et al.* Vitamin D status of clinical practice populations at higher latitudes: analysis and applications. *Int J Environ Res Public Health* 2009;**6**:151–73. doi:10.3390/ijerph6010151
148 Reeve J, Dowrick CF, Freeman GK, *et al.* Examining the practice of generalist expertise: a qualitative study identifying constraints and solutions. *JRSM Short Rep* 2013;**4**:2042533313510155. doi:10.1177/2042533313510155
149 Zwolsman SE, van Dijk N, de Waard MW. Observations of evidence-based medicine in general practice. *Perspect Med Educ* 2013;**2**:196–208. doi:10.1007/s40037-013-0078-8
150 Levine D, Bleakley A. Maximising medicine through aphorisms. Med. Educ. 2012;**46**:153–62. doi:10.1111/j.1365-2923.2011.04141.x
151 Woolf SH. The meaning of translational research and why it matters. *JAMA* 2008;**299**:211–3. doi:10.1001/jama.2007.26

152 WHO. Bridging the" know-do" gap: Meeting on Knowledge Translation in Global Health. Geneva: 2006. http://www.who.int/kms/WHO_EIP_KMS_2006_2.pdf (accessed 3 Oct 2013).
153 Crilly T, Jashapara A, Ferlie E. Research Utilisation & Knowledge Mobilisation: A Scoping Review of the Literature. 2010. http://www.nets.nihr.ac.uk/projects/hsdr/081801220
154 Greenhalgh T, Robert G, Bate P, *et al.* How to Spread Good Ideas A systematic review of the literature on diffusion , dissemination and sustainability of innovations in health service delivery and organisation. 2004.
155 Davies H, Nutley S, Walter I. Why 'knowledge transfer' is misconceived for applied social research. *J Health Serv Res Policy* 2008;**13**:188–90. doi:10.1258/jhsrp.2008.008055
156 Greenhalgh T, Wieringa S. Is it time to drop the 'knowledge translation' metaphor? A critical literature review. *J R Soc Med* 2011;**104**:501–9.
157 Nonaka I. A Dynamic Theory of Organizational Knowledge Creation. *Organ Sci* 1994;**5**:14–37.
158 Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. *BMC Med Res Methodol* 2009;**9**:59. doi:10.1186/1471-2288-9-59
159 Polanyi M. *The Tacit Dimension*. New York: : Doubleday & Company Inc 1966.
160 Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 7. Deciding what evidence to include. *Health Res Policy Syst* 2006;**4**:19. doi:10.1186/1478-4505-4-19
161 Ogden J, Bavalia K, Bull M, *et al.* 'I want more time with my doctor': a quantitative study of time and the consultation. *Fam Pract* 2004;**21**:479–83. doi:10.1093/fampra/cmh502
162 Heisenberg W. *Physics and philosophy*. New Ed. London: : Penguin Classics 2000.
163 James W. *The Principles of Psychology Vol 1*. Henry Holt New York 1890. https://archive.org/details/theprinciplesofp01jameuoft
164 Howick J, Glasziou P, Aronson JK. The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute? *J R Soc Med* 2009;**102**:186–94. doi:10.1258/jrsm.2009.090020
165 Clarke B, Gillies D, Illari P, *et al.* The evidence that evidence-based medicine omits. *Prev Med (Baltim)* 2012;**57**:745–7. doi:10.1016/j.ypmed.2012.10.020
166 Pearl J. *Causality: Models, Reasoning and Inference*. Cambridge university press 2009.
167 Tsoukas H. Do we really understand tacit knowledge? In: Easterby-Smith M, Lyles MA, eds. *The Blackwell Handbook of Organizational Knowledge Management*. Oxford: Blackwell. 2003. 410–27.
168 Bakhtin M. *Problems of Dostoevsky's poetics*. Theory and. Minneapolis: : University of Minnesota Press 1984.
169 Wittgenstein L. *Philosophical Investigations*. Oxford: : Wiley-Blackwell; 4th Edition edition 2009.
170 Timmermans S, Mauck A. The promises and pitfalls of evidence-based medicine. *Health Aff (Millwood)* 2005;**24**:18–28. doi:10.1377/hlthaff.24.1.18
171 Amin A, Roberts J. Knowing in action: Beyond communities of practice. *Res Policy* 2008;**37**:353–69. doi:10.1016/j.respol.2007.11.003
172 Contandriopoulos D, Lemire M, Denis J-L, *et al.* Knowledge exchange processes

in organizations and policy arenas: a narrative systematic review of the literature. *Milbank Q* 2010;**88**:444–83. doi:10.1111/j.1468-0009.2010.00608.x
173 Duggal R, Menkes DB. Evidence-based medicine in practice. *Int J Clin Pract* 2011;**65**:639–44. doi:10.1111/j.1742-1241.2011.02681.x
174 Oborn E, Dawson S. Knowledge and practice in multidisciplinary teams: Struggle, accommodation and privilege. *Hum Relations* 2010;**63**:1835–57. doi:10.1177/0018726710371237
175 Genuis SK, Genuis SJ. Exploring the continuum: medical information to effective clinical practice. Paper I: the translation of knowledge into clinical practice. *J Eval Clin Pract* 2006;**12**:49–62. doi:10.1111/j.1365-2753.2005.00608.x
176 Locke S. The Public Understanding of Science--A Rhetorical Invention. *Sci Technol Human Values* 2002;**27**:87–111. doi:10.1177/016224390202700104
177 Tuckett D. *Meetings between experts: an approach to sharing ideas in medical consultations*. London: : Tavistock Publications 1985.
178 Berwick DM. What 'patient-centered' should mean: confessions of an extremist. *Health Aff (Millwood)* 2009;**28**:w555-65. doi:10.1377/hlthaff.28.4.w555
179 Broekaert E, Autrique M, Vanderplasschen W, *et al.* 'The Human Prerogative': A Critical Analysis of Evidence-Based and Other Paradigms of Care in Substance Abuse Treatment. *Psychiatr Q* 2010;**81**:227–38. doi:10.1007/s11126-010-9132-4
180 Croskerry P. From Mindless to Mindful Practice — Cognitive Bias and Clinical Decision Making. *N Engl J Med* 2013;**368**:2445–8. doi:10.1056/NEJMp1303712
181 Mach E. *The analysis of sensations, and the relation of the physical to the psychical*. Digitizing. Chicago, London: : The Open Court Publishing Company 1914. https://archive.org/stream/analysisofsensat00mach/analysisofsensat00mach_djvu.txt
182 Tsoukas H. The tyranny of light. *Futures* 1997;**29**:827–43. doi:10.1016/S0016-3287(97)00035-9
183 Chang H. *Is Water H2O?* Dordrecht: : Springer Netherlands 2012. doi:10.1007/978-94-007-3932-1
184 Malterud K. The social construction of clinical knowledge - the context of culture and discourse. Commentary on Tonelli (2006), Integrating evidence into clinical practice: an alternative to evidence-based approaches. Journal of Evaluation in Clinical Practice 12, 24. *J Eval Clin Pract* 2006;**12**:292–5. doi:10.1111/j.1365-2753.2006.00591.x
185 Jonas WB. Scientific Evidence and Medical Practice The 'Drunkard's Walk'. *Arch Intern Med* 2009;**169**:649–50.
186 Valkenburg G, Achterhuis H, Nijhof A. Fundamental shortcomings of evidence-based medicine. *J Health Organ Manag* 2003;**17**:463–71. doi:10.1108/14777260310506614
187 Wood M, Ferlie E, Fitzgerald L. Achieving clinical behaviour change: a case of becoming indeterminate. 1998;**47**:1729–38.
188 Serres M. *The Parasite*. Baltimore and London: : The John Hopkins University Press 1982.
189 Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. *PLoS Med* 2010;**7**:1–5. doi:10.1371/journal.pmed.1000360

190 Catwell L, Sheikh A. Evaluating eHealth interventions: The need for continuous systemic evaluation. *PLoS Med* 2009;**6**:1–6. doi:10.1371/journal.pmed.1000126
191 Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. In: Gøtzsche PC, ed. *Cochrane Database of Systematic Reviews*. Chichester, UK: : John Wiley & Sons, Ltd 2013. CD001877. doi:10.1002/14651858.CD001877.pub5
192 Godlee F, Groves T. The new BMJ policy on sharing data from drug and. 2013;**7888**:1–3. doi:10.1136/bmj.e7888
193 Jefferson T, Doshi P. Multisystem failure: the story of anti-influenza drugs. *Bmj* 2014;**348**:g2263–g2263. doi:10.1136/bmj.g2263
194 Le Noury J, Nardo JM, Healy D, *et al.* Study 329 continuation phase: Safety and efficacy of paroxetine and imipramine in extended treatment of adolescent major depression. *Int J Risk Saf Med* 2016;**28**:143–61. doi:10.3233/JRS-160728
195 Green J. Epistemology , evidence and experience : evidence based health care in the work of Accident Alliances. *Sociol Health Illn* 2000;**22**:453–76.
196 Vandvik PO, Otto CM, Siemieniuk RA, *et al.* Transcatheter or surgical aortic valve replacement for patients with severe, symptomatic, aortic stenosis at low to intermediate surgical risk: a clinical practice guideline. *Bmj* 2016;**5085**:i5085. doi:10.1136/bmj.i5085
197 Schei E. Doctoring as leadership. The power to heal. *Perspect Biol Med* 2006;**49**:393–406.
198 Stolper E, Van De Wiel M, Van Royen P, *et al.* Gut feelings as a third track in general practitioners' diagnostic reasoning. *J Gen Intern Med* 2011;**26**:197–203. doi:10.1007/s11606-010-1524-5
199 Mmath KET, Zwarenstein M, Oxman AD, *et al.* Analysis A pragmatic – explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *Can Med Assoc J* 2009;**180**.
200 Greenhalgh T, Howick J, Maskrey N. Evidence based medicine: a movement in crisis? *BMJ Br Med J* 2014;**348**:g3725. doi:10.1136/bmj.g3725
201 Staley K, Doherty C. It ' s not evidence , it ' s insight : bringing patients ' perspectives into health technology appraisal at NICE. *Res Involv Engagem* 2016;:1–12. doi:10.1186/s40900-016-0018-y
202 Russell J, Greenhalgh T. Being 'rational' and being 'human': How National Health Service rationing decisions are constructed as rational by resource allocation panels. *Health (London)* 2013;**18**:441–57. doi:10.1177/1363459313507586
203 Jonsen AR, Toulmin S. *The Abuse of Casuistry*. University of California Press; Reprint edition (1 July 1992) 1992.
204 Kerry R, Eriksen TE, Lie SAN, *et al.* Causation and evidence-based practice: An ontological review. *J Eval Clin Pract* 2012;**18**:1006–12. doi:10.1111/j.1365-2753.2012.01908.x
205 Askheim C, Sandset T, Engebretsen E. Who cares? The lost legacy of Archie Cochrane. *Med Humanit* 2017;**43**:41–6. doi:10.1136/medhum-2016-011037
206 Zuiderent-Jerak T, Jerak-Zuiderent S, Bovenkamp H, *et al.* Variatie in Richtlijnen; wat is het probleem. Rotterdam: 2011. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Variatie+in+Richtlijnen+Wat+is+het+probleem+?#0
207 van de Bovenkamp HM, Zuiderent-Jerak T. An empirical study of patient participation in guideline development: Exploring the potential for articulating

patient knowledge in evidence-based epistemic settings. *Heal Expect* 2015;**18**:942–55. doi:10.1111/hex.12067
208 Wennberg JE. Dealing with medical practice variations: A proposal for action. *Health Aff* 1984;**3**:6–32. doi:10.1377/hlthaff.3.2.6
209 Walach H, Falkenberg T, F?nneb? V, *et al.* Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol* 2006;**6**. doi:10.1186/1471-2288-6-29
210 Harder T, Abu Sin M, Bosch-Capblanch X, *et al.* Towards a framework for evaluating and grading evidence in public health. *Health Policy (New York)* 2015;**119**:732–6. doi:10.1016/j.healthpol.2015.02.010
211 Bradford-Hill A. The Enviroment and Disease: Association or Causation? *Proc R Soc Med* 1965;**58**:295–300. doi:DOI: 10.1016/j.tourman.2009.12.005
212 Illari PM, Williamson J. What is a mechanism? Thinking about mechanisms across the sciences. *Eur J Philos Sci* 2012;**2**:119–35. doi:10.1007/s13194-011-0038-2
213 van Dijk FJH, Verbeek JH, Hoving JL, *et al.* A knowledge infrastructure for occupational safety and health. *J Occup Environ Med* 2010;**52**:1262–8. doi:10.1097/JOM.0b013e318202f2c5
214 Lewin S, Glenton C, Munthe-Kaas H, *et al.* Using Qualitative Evidence in Decision Making for Health and Social Interventions: An Approach to Assess Confidence in Findings from Qualitative Evidence Syntheses (GRADE-CERQual). *PLoS Med* 2015;**12**:1–18. doi:10.1371/journal.pmed.1001895
215 Ring N, Ritchie K, Mandava L, *et al. A guide to synthesising qualitative research for researchers undertaking health technology assessments and systematic reviews*. NHS Quality Improvement Scotland 2011. http://www.iiip.scot.nhs.uk/media/CLT/ResourceUploads/1011201/HTA_MethodsofSynthesisingQualitativeLiterature_DEC101.pdf%5Cnhttps://dspace.stir.ac.uk/handle/1893/3205
216 Juritzen TI, Soberg HL, Roe C, *et al.* The One or the Many: Quantified Subjectivity and Aggregated Uniqueness in Qualitative Rehabilitation Research. *Qual Health Res* Published Online First: 2016. doi:10.1177/1049732316668297
217 Moberg J, Alonso-Coello P, Oxman A. GRADE Evidence to Decision (EtD) Frameworks Guidance. Version 1.1. GRADE Work. Gr. 2015.https://ietd.epistemonikos.org/#/help/guidance
218 Verkerk K, Van Veenendaal H, Severens JL, *et al.* Considered judgement in evidence-based guideline development. *Int J Qual Heal Care* 2006;**18**:365–9. doi:10.1093/intqhc/mzl040
219 Calderón C, Rotaeche R, Etxebarria A, *et al.* Gaining insight into the Clinical Practice Guideline development processes: qualitative study in a workshop to implement the {GRADE} proposal in Spain. *{BMC} Heal Serv Res* 2006;**6**. doi:10.1186/1472-6963-6-138
220 Lomas J. Words Without Action? The Production, Dissemination, and Impact of Consensus Recommendations. *Annu Rev Public Health* 1991;:41–65.
221 Douven I. Abduction. 2011.http://plato.stanford.edu/entries/abduction/#AbdVerBayConThe (accessed 23 Dec 2015).
222 Kriebel D, Tickner J, Epstein P, *et al.* The precautionary principle in

environmental science. *Environ Health Perspect* 2001;**109**:871–6.
223 Pollock JL. The Logical Foundations of Means-End Reasoning. In: *Common Sense, Reasoning, and Rationality*. Oxford University Press 2002. 60–77. doi:10.1093/0195147669.003.0003
224 Stanford University., Center for the Study of Language and Information (U.S.). Truth. Stanford encyclopedia of philosophy. 2013.https://plato.stanford.edu/entries/truth/ (accessed 1 Apr 2018).
225 Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health* 2004;**58**:635–41. doi:10.1136/jech.2003.008466
226 Grimes DA, Schulz KF. Bias and ausal associations in observational research. *Lancet* 2002;**359**:248–52. doi:http://dx.doi.org/10.1016/S0140-6736(02)07451-2
227 Paradis C. Bias in surgical research. *Ann Surg* 2008;**248**:180–8. doi:10.1097/SLA.0b013e318176bf4b
228 Dohoo IR. Bias-Is it a problem, and what should we do? *Prev Vet Med* 2014;**113**:331–7. doi:10.1016/j.prevetmed.2013.10.008
229 Song F, Parekh S, Hooper L, *et al.* Dissemination and publication of research findings: An updated review of related biases. *Health Technol Assess (Rockv)* 2010;**14**:1–220. doi:10.3310/hta14080
230 Wunsch H, Linde-Zwirble WT, C. Angus D. Methods to adjust for bias and confounding in critical care health services research involving observational data. *J Crit Care* 2006;**21**:1–7. doi:10.1016/j.jcrc.2006.01.004
231 Pannucci CJ, Wilkins EG. Identifying and Avoiding Bias in Research. *Plast Reconstr Surg* 2010;**126**:619–25. doi:10.1097/PRS.0b013e3181de24bc
232 Catalogue of Bias - Catalog of Bias. https://catalogofbias.org/ (accessed 25 Feb 2018).
233 D'Ancona M. *Post-Truth: the New War on Truth and How to Fight Back*. 2017.
234 Davis E. *Post-Truth: Why We Have Reached Peak Bullshit and What We Can Do About It*. Little, Brown 2017. https://books.google.co.uk/books/about/Post_Truth.html?id=0ejunQEACAAJ&redir_esc=y (accessed 25 Feb 2018).
235 Neiman S. *Widerstand der Vernunft: Ein Manifest in postfaktischen Zeiten*. Ecowin 2017. https://books.google.co.uk/books/about/Widerstand_der_Vernunft.html?id=iNgkDwAAQBAJ&redir_esc=y (accessed 25 Feb 2018).
236 Sismondo S. Post-truth? *Soc Stud Sci* 2017;**47**:3–6. doi:10.1177/0306312717692076
237 Lewandowsky S, Ecker UKH, Cook J. Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *J Appl Res Mem Cogn* 2017;**6**:353–69. doi:10.1016/j.jarmac.2017.07.008
238 Stengers I. *Sciences et pouvoirs. La democratie face a la technoscience*. La Decouverte / Poche 2002.
239 Blackburn S. *Truth: A guide for the perplexed*. Penguin 2006.
240 Burgess AG, Burgess JP. *Truth*. Princeton: : Princeton University Press 2014. doi:10.1515/9781400838691
241 Truth | Internet Encyclopedia of Philosophy. http://www.iep.utm.edu/truth/#H3 (accessed 12 Nov 2014).

242 Caputo J. *Truth*. Penguin 2013.
243 Sismondo S. *Science without Myth*. New York: : State University of New York 1996.
244 Aldemer R. Peirce ' s Thirteen Theories of Truth. *Trans Charles S Peirce Soc* 1985;**21**:77–94.http://www.jstor.org/stable/40320077
245 Peirce CS. Truth and Falsity. In: Baldwin JM, ed. *Dictionary of Philosophy and Psychology*. London: : MacMillan and co. 1902. 716–20.
246 Fisher SRA. The Design of Experiments. Hafner Publ. Co. 1971;:1–27. doi:n/a
247 Hill AB. Principles of Medical Statistics. *Lancet* 1937;**January 2**:41–3.
248 Armitage P. Fisher, Bradford Hill, and randomization. *Int J Epidemiol* 2003;**32**:925–8. doi:10.1093/ije/dyg286
249 Cochrane AL. *Effectiveness and Efficiency: Random Reflections on Health Services*. 1972.
250 Kahneman D, Sugden R. Experienced Utility as a Standard of Policy Evaluation. *Environ Resour Econ* 2005;**32**:161–81. doi:10.1007/s10640-005-6032-4
251 Lambert Mogiliansky A, Zamir S, Zwirn H. Type indeterminacy: A model of the KT(Kahneman-Tversky)-man. *J Math Psychol* 2009;**53**:349–61. doi:10.1016/j.jmp.2009.01.001
252 Norman GR, Eva KW. Diagnostic error and clinical reasoning. *Med Educ* 2010;**44**:94–100. doi:10.1111/j.1365-2923.2009.03507.x
253 Nussbaum M. *Upheavals of Thought: the Intelligence of Emotions.* Cambridge, MA: : Cambridge University Press 2001.
254 Foucault M. *Discipline and punish : the birth of the prison*. Penguin Books 1979.
255 Wendt A. *Quantum Mind and Social Science. Unifying Phisical and Social Ontology*. 2015. doi:10.1017/CBO9781316005163
256 Sackett DL. Equipoise, a term whose time (if it ever came) has surely gone. *CMAJ* 2000;**163**:835–6.
257 Anjum RL. Evidence Based or Person Centered? An Ontological Debate. *Eur J Pers Centered Healthc* 2016;**4**:421–9.
258 Papineau D. The Evolution of Means-End Reasoning. 2017;:145–78.
259 Russo F, Williamson J. Interpreting Causality in the Health Sciences. *Int Stud Philos Sci* 2007;**21**:157–70. doi:10.1080/02698590701498084
260 Dragulinescu S. Mechanisms and Difference-Making. *Acta Anal* 2017;**32**:29–54. doi:10.1007/s12136-016-0292-1
261 Eva KW, Hatala RM, LeBlanc VR, *et al.* Teaching from the clinical reasoning literature: Combined reasoning strategies help novice diagnosticians overcome misleading information. *Med Educ* 2007;**41**:1152–8. doi:10.1111/j.1365-2923.2007.02923.x
262 Graber M. Metacognitive training to reduce diagnostic errors: ready for prime time? *Acad Med* 2003;**78**:781. doi:10.1097/00001888-200308000-00004
263 Graber ML, Franklin N, Gordon R. DIagnostic error in internal medicine. *Arch Intern Med* 2005;**165**:1493–9. doi:10.1001/archinte.165.13.1493
264 Russell J, Greenhalgh T, Byrne E, *et al.* Recognizing rhetoric in health care policy analysis. *J Health Serv Res Policy* 2008;**13**:40–6. doi:10.1258/jhsrp.2007.006029
265 Greenhalgh T, Russell J. Reframing evidence synthesis as rhetorical action in the policy making drama. *Healthc Policy* 2006;**1**:34–42.
266 Public Health England. Scarlet fever: guidance and data.

2018.https://www.gov.uk/government/collections/scarlet-fever-guidance-and-data
267 NHG. NHG-Standaard Acute keelpijn. https://www.nhg.org/standaarden/volledig/nhg-standaard-acute-keelpijn (accessed 6 Apr 2018).
268 Spinks A, Glasziou P, Del Mar C. Antibiotics for sore throat (Review) SUMMARY OF FINDINGS FOR THE MAIN COMPARISON. *Cochrane Database Syst Rev* Published Online First: 2013. doi:10.1002/14651858.CD000023.pub4.www.cochranelibrary.com
269 Aristotle. *Physics*. Translated. The Internet Classics Archive http://classics.mit.edu//Aristotle/physics.html
270 Nussbaum M. *The Fragility of Goodness*. Cambridge: : Cambridge university press 2001.
271 Cassam Q. Vice epistemology. *Monist* 2016;**99**:159–80. doi:10.1093/monist/onv034
272 Kant I. *Critique of Pure Reason*. The Cambri. Cambridge university press 1998. http://www.nature.com/articles/125557a0
273 Kahneman D. *Thinking, fast and slow.* Penguin 2013.
274 Hurwitz B. How does evidence based guidance influence determination of medical negligence? *BMJ* 2004;**329**:1024–8. doi:http://dx.doi.org/10.1136/bmj.329.7473.1024
275 James W. *Pragmatism and The Meaning of Truth*. Rough Draft Printing 2011.
276 Yeh B, Eskin B. Should sore throats be treated with antibiotics? *Ann Emerg Med* 2005;**45**:82–4. doi:10.1016/j.annemergmed.2004.08.010
277 Hocking WE. *The meaning of God in human experience*. New Haven and London: : Yale University Press 1912.
278 NICE. CG71 Familial hypercholesterolaemia: NICE guideline. https://www.nice.org.uk/guidance/cg71 (accessed 8 Apr 2014).
279 Wyatt JC. Management of explicit and tacit knowledge. *J R Soc Med* 2001;**94**:6–9.
280 Nonaka I, von Krogh G. Perspective--Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory. *Organ Sci* 2009;**20**:635–52. doi:10.1287/orsc.1080.0412
281 Sujan M. An organisation without a memory: A qualitative study of hospital staff perceptions on reporting and organisational learning for patient safety. *Reliab Eng Syst Saf* 2015;**144**:45–52. doi:10.1016/j.ress.2015.07.011
282 Brown JS, Duguid P. Organizational Learning And Communities Of-Practice : Toward A Unified View Of Working , Learning , And Innovation. *Organ Sci* 1991;**2**:40–57. doi:10.1287/orsc.2.1.40
283 Gabbay J, Le May A. Mindlines: Making sense of evidence in practice. *Br J Gen Pract* 2016;**66**:402–3. doi:10.3399/bjgp16X686221
284 May L. *The socially responsive self : social theory and professional ethics*. University of Chicago Press 1996.
285 Chang H. *Is Water H2O?* Dordrecht: : Springer Netherlands 2012. doi:10.1007/978-94-007-3932-1
286 Hunter K. 'Don't think zebras': Uncertainty, interpretation, and the place of paradox in clinical education. *Theor Med* 1996;**17**:225–41.

doi:10.1007/BF00489447
287 Greenhalgh T, Hurwitz B. Narrative based medicine: why study narrative? *BMJ* 1999;**318**:48–50. doi:10.1136/bmj.318.7175.48
288 Launer J. Narrative-based medicine: A passing fad or a giant leap for general practice? *Br J Gen Pract* 2003;**53**:91–2.
289 Kalitzkus V, Matthiessen PF. Narrative-based medicine: potential, pitfalls, and practice. *Perm J* 2009;**13**:80–6.
290 Balint M. The doctor, his patient, and the illness. *Lancet* 1955;**268**:683–8. doi:10.1136/bmj.39384.467928.94
291 Weick K. Sensemaking in Organizations: Small Structures with Large Consequences. In: Keith Murnighan J, ed. *Social Psychology in Organizations*. New Jersey: 1993. 10–37.
292 James W. *Pragmatism: A New Name for Some Old Ways of Thinking*. The Projec. Cambridge: : Havard University Press 1907. http://livros01.livrosgratis.com.br/gu005116.pdf
293 Wieringa S, Engebretsen E, Heggen K, *et al.* Can better clinical practice guidelines help reduce epistemic uncertainty? An ethnographic study of guideline development panels in three countries. *Acad Med* 2019.
294 Treweek S, Oxman AD, Alderson P, *et al.* Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence (DECIDE): protocol and preliminary results. *Implement Sci* 2013;**8**:1–12. doi:10.1186/1748-5908-8-6
295 Knaapen L. Being 'evidence-based' in the absence of evidence: The management of non-evidence in guideline development. *Soc Stud Sci* 2013;**43**:681–706. doi:10.1177/0306312713483679
296 Wieringa S, Engebretsen E, Heggen K, *et al.* Has evidence-based medicine ever been modern? A Latour-inspired understanding of a changing EBM. *J Eval Clin Pract* 2017;:1–7. doi:10.1111/jep.12752
297 Greenhalgh T, Wieringa S. Is it time to drop the 'knowledge translation' metaphor? A critical literature review. *J R Soc Med* 2011;**104**:501–9. doi:10.1258/jrsm.2011.110285
298 Wyer P, Alves S. 'All the King ' s horses . . .': the problematical fate of born-again evidence-based medicine : commentary on Greenhalgh , T ., Snow , R ., Ryan , S ., Rees , S ., and Salisbury , H . (2015) six ' biases ' against patients and carers in evidence-based m. *J Eval Clin Pract* 2015;**21**.
299 Sur R, Dahm P. History of evidence-based medicine. *Indian J Urol* 2012;**27**:487. doi:10.4103/0970-1591.91438
300 Gabbay J, Le May A. Transforming knowledge: the theoretical and practical implications of 'mindlines' - Oslo Workshop 'The foundations of evidence-based medicine' 2016. 2016.
301 Zwolsman SE, van Dijk N, de Waard MW. Observations of evidence-based medicine in general practice. *Perspect Med Educ* 2013;**2**:196–208. doi:10.1007/s40037-013-0078-8
302 Zwolsman S, Hooft L, Waard MW, *et al.* medicine : 2012;:511–21. doi:10.3399/bjgp12X652382.e511
303 May M. Twenty-five ways clinical trials have changed in. *Nat Med* 2019;**25**:2–5. doi:10.1038/s41591-018-0314-1

304 Hurst D, Greenhalgh T. Knowing in general dental practice: Anticipation, constraint, and collective bricolage. *J Eval Clin Pract* 2018;:1–9. doi:10.1111/jep.13051
305 Beckett K, Farr M, Kothari A, *et al.* Embracing complexity and uncertainty to create impact: Exploring the processes and transformative potential of co-produced research through development of a social impact model. *Heal Res Policy Syst* 2018;**16**:1–18. doi:10.1186/s12961-018-0375-0
306 Aerts D, Beltran L, Geriente S, *et al.* Quantum Theory Methods as a Possible Alternative for the Double-Blind Gold Standard of Evidence-Based Medicine: Outlining a New Research Program. *Found Sci* 2018;**24**:217–25. doi:10.1007/s10699-018-9572-0
307 Reynen E, Robson R, Ivory J, *et al.* A retrospective comparison of systematic reviews with same-topic rapid reviews. *J Clin Epidemiol* 2018;**96**:23–34. doi:10.1016/j.jclinepi.2017.12.001
308 Shlonsky A, Mildon R. Methodological pluralism in the age of evidence-informed practice and policy. *Scand J Public Health* 2014;**42**:18–27. doi:10.1177/1403494813516716
309 Parkkinen V-P, Wallmann C, Wilde M, *et al.* Evaluating evidence of mechanisms in medicine: principles and procedures. 2018. http://ebmplus.org/about/papers-articles/
310 Greenhalgh T, Annandale E, Ashcroft R, *et al.* An open letter to the BMJ editors on qualitative research. *BMJ* 2016;**352**. doi:10.1136/bmj.i563
311 Rowe R, McDaid D. Implementation: the need for a contextual approach to the implementation of musculoskeletal guidelines. *Best Pract Res Clin Rheumatol* 2007;**21**:205–19. doi:10.1016/j.berh.2006.08.002
312 The Council for Health and Society. No evidence without context About the illusion of evidence - based practice in healthcare. The Hague: 2017. https://www.raadrvs.nl/en/item/no-evidence-without-context
313 Elwyn G, Wieringa S, Greenhalgh T. Clinical encounters in the post-guidelines era. *Bmj* 2016;**3200**:i3200. doi:10.1136/bmj.i3200
314 Elwyn G, Pickles T, Edwards A, *et al.* Supporting shared decision making using an Option Grid for osteoarthritis of the knee in an interface musculoskeletal clinic: A stepped wedge trial. *Patient Educ Couns* 2016;**99**:571–7. doi:10.1016/j.pec.2015.10.011
315 Wyatt KD, Branda ME, Anderson RT, *et al.* Peering into the black box: A meta-analysis of how clinicians use decision aids during clinical encounters. *Implement Sci* 2014;**9**:1–10. doi:10.1186/1748-5908-9-26
316 Hine C. Internet Research and the Sociology of Cyber-Social-Scientific Knowledge. *Inf Soc* 2005;**21**:239–48. doi:10.1080/01972240591007553
317 Tonelli MR, Upshur REG. A Philosophical Approach to Addressing Uncertainty in Medical Education. *Acad Med* 2018;:1. doi:10.1097/ACM.0000000000002512
318 Low M. A novel clinical framework: The use of dispositions in clinical practice. A person centred approach. *J Eval Clin Pract* 2017;**23**:1062–70. doi:10.1111/jep.12713
319 Wendt A. The mind – body problem and social science : Motivating a quantum social theory. *J Theory Soc Behav* 2018;:188–204. doi:10.1111/jtsb.12171
320 Fraser SW, Greenhalgh T. Coping with complexity: educating for capability. *BMJ*

2001;**323**:799–803.
321 Medicine in the digital age. *Nat Med* 2019;**25**:1. doi:10.1038/s41591-018-0322-1
322 Safran C. Using routinely collected data for clinical research. *Stat Med* 1991;**10**:559–64.
323 Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *Can Med Assoc J* 2016;**188**:E158–64. doi:10.1503/cmaj.150653
324 Agha RA, Fowler AJ, Saeta A, *et al.* The SCARE Statement: Consensus-based surgical case report guidelines. *Int J Surg* 2016;**34**:180–6. doi:10.1016/j.ijsu.2016.08.014
325 Vandenbroucke JP. In Defense of Case Reports and Case Series. *Ann Intern Med* 2001;:330–4.
326 Pedersen N, And JLL, Wright C. Pluralist Theories of Truth. Stanford Encycl. Philos. (Winter 2018 Ed. Edward N. Zalta. 2018.https://plato.stanford.edu/archives/win2018/entries/truth-pluralist/