Valuing and Evaluating Evidence in Medicine

by

Kirstin Borgerson

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Philosophy
University of Toronto

© Copyright by Kirstin Borgerson (2008)
Medical decisions should be based on good evidence. But this does not mean that health care professionals should practice evidence-based medicine. This dissertation explores how these two positions come apart, why they come apart, and what we should do about it. I begin by answering the descriptive question, what are current standards of evidence in medicine? I then provide a detailed critique of these standards. Finally, I address the more difficult normative question, how should we determine standards of evidence in medicine?

In medicine, standards of evidence have been established by the pervasive evidence-based medicine (EBM) movement. Until now, these standards have not been subjected to comprehensive philosophical scrutiny. I outline and defend a theory of knowledge – a version of Helen Longino’s Critical Contextual Empiricism (CCE) – which enables me to critically evaluate EBM. My version of CCE emphasizes the critical evaluation of background assumptions. In accordance with this, I identify and critically evaluate the three substantive assumptions underlying EBM. First, I argue that medicine should not be held to the restrictive definition of science assumed by proponents of EBM. Second, I argue that epidemiological evidence should not be the only “base” of medical decisions. Third, I argue that not only is the particular hierarchy of evidence assumed by EBM unjustified, but that any attempt to hierarchically rank
research methods is incoherent and unjustifiably restricts medical knowledge. Current standards of evidence divert attention from many legitimate sources of evidence. This distorts medical research and practice.

In the remainder of the dissertation I propose means for improving not only current standards of medical evidence but also the process of producing and defending future standards. On the basis of four CCE norms, I argue that we have reason to protect and promote those features of the medical community that facilitate diversity, transparency, and critical interaction. Only then can we ensure that the medical community retains its ability to produce evidence that is both rigorous and relevant to practice.
Acknowledgements

I wish to thank my supervisor, Margaret Morrison, for her good judgment on everything from the scope and content of the thesis to the inner workings of academia. Thanks also to Ross Upshur for his good humour, love of philosophy and heretical insider perspective on EBM, and Barry Brown for sharing his wealth of experience on everything related to alternative health care. In the course of a PhD in philosophy you expect to learn a lot about your topic of research. In addition you hope that you will learn about what it means to be a philosopher and what value philosophy brings to the world. I learned the most about these deeply important issues from my graduate student colleagues. Particular thanks to Joseph Millum, Michael Garnett, Danielle Bromwich, Marika Warren, Vida Panitch and Lauren Bialystok.

I would also like to thank my first philosophy professor, Rudy Krutzen, for his persistence; Kathryn Morgan for advocating for me when it really mattered; Heather Boon for her encouragement; Jim Brown for his mentorship; Bob Perlman from Perspectives in Biology and Medicine for believing that two graduate students in philosophy had something important to say and for giving us a venue to say it; Joshua Goldstein for his unwavering support; the EBM infidels, Robyn Bluhm and Maya Goldenberg, who have become not only colleagues and collaborators, but also friends; fellow members of the Varsity Rowing Team – particularly Kiran van Rijn – for sharing sunrises over Lake Ontario; Shannon Urbaniak for setting such a fine example in so many ways; and Murray Enkin for reminding me that there are some advances that have to be made one step at a time, whether in medicine or in life. Thanks also to Joanne Beckett and Dave White, Donna and Neil Larsen, Pat and Trent Cooley, and all the Svenkes.

I was generously supported, both financially and intellectually, by several programs while at the University of Toronto. I received funding from the Ontario Graduate Scholarship program on several occasions. In addition, the Comparative Program on Health and Society (CPHS) awarded me two doctoral fellowships and one research associateship and provided me with opportunities to give my first public seminar presentation and to publish my first working paper. I am deeply grateful to everyone at the CPHS for their support. I was also awarded a CIHR Strategic
Research and Training Doctoral Fellowship in Health Care, Technology and Place in the last two years of my PhD. Through this program I met a number of researchers from other disciplines, many of whom introduced me to new areas of research. I am confident that these encounters will shape my future research for the better.

Finally, my deepest thanks go to my family. Thanks to my grandparents, two of whom experienced the worst of health care, and two of whom experienced the best. I hope that my research goes some way to shifting those odds for future patients. Thanks to my mother, Val Drummond, whose intelligence is equaled only by her compassion. I am completing a PhD today because she is such a strong feminist role model and because she has provided me with unconditional support at every stage of my life. Sharing equal credit for phenomenal acts of parenting is my endlessly energetic and incredibly generous father, Lon Borgerson, who always knew I was bright but who pushed me to be more than that: to be creative, to be passionate, and to make a difference. Thanks to my sister, Erika Borgerson, for being my best friend for as long as I can remember and for being there every time I needed to talk. Finally, thanks to my partner, Kieran Cooley, for so many things but most importantly because he always knows how to make me smile.
Contents
Valuing and Evaluating Evidence in Medicine

Abstract ii
Acknowledgements iv
Contents vi

Introduction 1
1. Why Standards Matter: The ECMO Case 1
2. Standards of Evidence and the Rise of Evidence-based Medicine 4
3. Locating the Project within Philosophy 5
   3.1. The Role of Philosophy
   3.2. Bioethics and Philosophy of Medicine
   3.3. Epistemological Framework
   3.4. Context
4. The Argument Ahead 11
   4.1. Chapter One: Social Epistemology
   4.2. Chapter Two: Critical Contextual Empiricism
   4.3. Chapter Three: Evidence-based Medicine
   4.4. Chapter Four: Critiquing the Background Assumptions of Evidence-based Medicine: Part I
   4.5. Chapter Five: Critiquing the Background Assumptions of Evidence-based Medicine: Part II
   4.7. Chapter Seven: The Future of Evidence-based Medicine: Interactive Objectivity and the Open Science Movement
5. Aim of the Dissertation 15

Chapter One
Social Epistemology 17
1. Individual and Social Epistemology 17
2. Social Epistemology 18
   2.1. Past and Present
   2.2. Reasons for Social Epistemology
3. Criteria for Choosing a Theory Within Social Epistemology 23

4. Applying the Criteria 26
   4.1. Focus on Epistemic Justification  
   4.2. Preservation of Individualistic Intuitions  
      4.2.1. Collective Entities  
      4.2.2. Inter-individual Relationships  
   4.3. Mechanisms for Managing Social Values  
   4.4. Robust Relationship between the Rational and the Social  
      4.4.1. Observation  
      4.4.2. Reason  

5. Critical Contextual Empiricism 34
   5.1. Naturalized Epistemology  
   5.2. Feminist Epistemology  

6. Summary 38

**Chapter Two**  
**Critical Contextual Empiricism** 39

1. Constitutive and Contextual Values 40
2. Background Assumptions 41
3. CCE Norms 44
4. Modifications to CCE 45
5. Cultivating Diverse Perspectives 48
   5.1. Preliminaries  
   5.2. Standard Epistemological Argument for Diversity  
   5.3. CCE Argument for Diversity  
6. Shared Public Standards 54
   6.1. The Tension  
   6.2. CCE and Community Standards  
   6.3. Local Community Standards and Regulatory Standards  
   6.4. Requirements for Inter-community Debate  
   6.5. Conclusions  
7. Clarification: Three Senses of Knowledge 61
   7.1. Agency: Who Knows?  
   7.2. Content: What Do They Know?  
   7.3. Warrant: How Do They Prove That They Know?  
   7.4. The Traditional Triumvirate  
   7.5. Avoiding Misunderstandings  
8. Objections and Replies 65
   8.1. Failure to Engage with Messy “Real” World of Science
1. Should Medicine Be More Scientific? 95
   1.1. Medical Research
      1.1.1. The Received View of Science
      1.1.2. EBM and Science
   1.2. Medical Practice
      1.2.1. Patient Outcomes
      1.2.2. Individualized Care
      1.2.3. Shared Decision-making

2. Should Clinical Research Evidence Form the Basis of Medical Decision-making? 108
   2.1. Not Authority?
   2.2. Not Pathophysiology?
      2.2.1. Relationship Between Bench Science and Clinical Research
         2.2.1.1. Homeopathy
         2.2.1.2. Naïve Empiricism
         2.2.1.3. Theoretical Framework
   2.3. Why a Base?

3. Summary 116

Chapter Five
Critiquing the Background Assumptions of Evidence-based Medicine: Part II 118

1. Precision 119
   1.1. Probability
      1.1.1. Hypothesis Testing
   1.2. Meta-analyses and EBM
   1.3. Critique
   1.4. Statistical Style of Reasoning
   1.5. Conclusions on Precision

2. Randomization 129
   2.1. Randomized Trials and EBM
   2.2. Non-randomized Trials Exaggerate Treatment Effects
   2.3. Non-randomized Trials Are Misleading
   2.4. Non-randomized Trials Cannot Balance Groups
      2.4.1. Aside: Table 1
   2.5. Non-randomized Trials Identify Correlations (Not Causes)
      2.5.1. Aside: Side-effects
   2.6. Conclusions on Randomization

3. Managing Bias 140
   3.1. Bias and EBM
   3.2. Catalogues of Bias
   3.3. Bias in RCTs
   3.4. Conclusions on Bias

4. Quantifiability 147
9.1.2.3. Preference Trials
9.1.2.4. N-of-1 Trials
9.1.3. Appropriate Outcome Measures: MYMOP
9.1.4. Data Analysis
9.1.5. Observational Research
9.1.6. Qualitative Research

10. Lessons from CAM
194

11. Critical Analysis
195
  11.1. Objection
  11.2. Reply

12. Nursing
198

13. Summary
200

Chapter Seven
The Future of Evidence-based Medicine: Interactive Objectivity and the Open Science Movement
201

1. A Short History of Objectivity in Science
202

2. Catalogue of Objectivity
204
  2.1. Experimental Objectivity
  2.2. Individual Objectivity
  2.3. Social Objectivity
    2.3.1. Concordant Objectivity
    2.3.2. Procedural Objectivity
    2.3.3. Interactive Objectivity

3. CCE Account of Objectivity
210
  3.1. Experimental Objectivity
  3.2. Individual Objectivity
  3.3. Social Objectivity
    3.3.1. Concordant Objectivity
    3.3.2. Procedural Objectivity
    3.3.3. Interactive Objectivity

4. Concerns
216

5. Objectivity and Values in Medical Research
218
  5.1. Why Procedural Objectivity?
  5.2. The Procedural Objectivity of RCTs
  5.3. The Evaluation-based Education Analogy
  5.4. Economic Models and RCTs
  5.5. Method and Methods

6. Interactive Objectivity and Medicine
224
  6.1. Recognized Avenues for Criticism: The Open Science Movement
6.1.1. Regulation of Research Contracts
6.1.2. Declaration of Competing Interests
6.1.3. Prohibition on Ghost Authorships
6.1.4. Clinical Trials Registry
6.1.5. Open Peer-review
6.1.6. Open Access Journals
6.1.7. Open Submission Process
6.1.8. Public Funding of Research
6.1.9. Emphasis on Original Research
6.1.10. Medical Teams
6.1.11. Analytic Training in Medical Education

7. Summary 235

Conclusion 237

1. Remaining Problems and Future Research 238
   1.1. Uncertainty
   1.2. Quantity of Publications
   1.3. Tendency to Prefer Rules over Reasoning
   1.4. Ethics and Epistemology

2. Themes 241

3. Final Words 241

Bibliography 245
Introduction

Medical decisions should be based on good evidence. But this does not mean that health care professionals should practice evidence-based medicine (EBM). This dissertation explores how these two positions come apart, why they come apart, and what we should do about it. In the chapters that follow, I identify the assumptions underlying current standards of medical evidence, evaluate those assumptions, and suggest structural changes that could enhance the capacity of medical research communities to produce knowledge. In this introduction I motivate the project, situate it within philosophy, and provide an overview of the chapters to come.

1. Why Standards Matter: ECMO

Standards of evidence have a direct impact on the lives and deaths of human beings. The following case demonstrates this impact unambiguously.\(^1\) Persistent pulmonary hypertension (PPHS) is a condition that affects neonates (newborns), and until the late 1970s PPHS had an established mortality rate of around 80%. Neonates with PPHS have underdeveloped lungs and as a result are unable to adequately oxygenate their blood. Because the etiology of the disease was well known, researchers were able to develop a treatment – called “extracorporeal membrane oxygenation” (ECMO) which involves removing blood before it reaches the lungs, oxygenating it artificially with a machine, warming it back to body temperature, and then returning it to the body. In effect, what the treatment does is keep the infant alive while the lungs have a chance to complete their development. The initial results were impressive. 45 neonates were treated over a period of several years and mortality rates dropped below 20% for that group.\(^2\) This dramatic improvement in survival was not linked to any other suspected confounding factors and the population (neonates) was hardly in a position to influence the data through expectations of success of a new treatment or the like. Because all infants were treated

---

1 The ECMO case was brought to my attention by Robyn Bluhm, and I have since benefited from the description and analysis offered by John Worrall in his article “Why There’s No Cause to Randomize.” No doubt there are many other cases in which unnecessary research, conducted out of a misguided commitment to inappropriate standards of scientific rigor in medical research, resulted in unjustified pain, suffering, disease and death. The ECMO case, however, does a particularly good job of highlighting what goes wrong when standards of evidence are applied uncritically. There are no clear distinctions between ethical and epistemological issues in this domain.

2 Bartlett et al., “Extracorporeal Membrane Oxygenation for Newborn Respiratory Failure.”
with ECMO, there were none of the standard concerns with selection bias, and it was extremely unlikely that the disease had radically changed in character in that short time period. No plausible alternative explanations for the change in survival rate were offered.\(^3\)

In spite of the dramatic results of the new treatment, researchers felt compelled to perform a randomized controlled trial (RCT), which tests new treatments against placebo or best available treatment.\(^4\) Though they fully expected infants in the ECMO group to live and those in the control group to die it was decided that only an RCT could establish their findings conclusively. Until then all they had was the historically-controlled data from their initial investigations. This data, when analyzed, amounted to a case-control study, which is not regarded as especially good evidence in medicine.\(^5\) There was a particular concern that fellow medical practitioners would dismiss any results that fell short of the RCT standard.\(^6\) So they went ahead with an RCT, the results of which were published in 1985.\(^7\) The researchers used a modified randomization protocol that allowed them to increase the odds of receiving ECMO with each positive result. Because of this methodology – proposed, no doubt, to ease the consciences of those participating – 12 infants were enrolled in the trial, 11 of whom received ECMO and lived and 1 who received standard care and died. Response by the medical community indicated that these results were not convincing because the control group consisted of only one patient. Critics called for further research.\(^8\) A second randomized trial was performed.

The results of the second, fully randomized trial enrolled 19 infants. 9 received ECMO and lived and of the 10 receiving standard care 6 survived.\(^9\) The insistence that a second RCT be conducted directly contributed to the deaths of 4 infants given standard care. The status of the evidence produced by the historically controlled initial data, the first partially-randomized trial and the second fully-randomized trial is clearly at issue here. If the standards of evidence accepted by the medical community had allowed the initial historically controlled results to accept...
count as significant evidence of efficacy, the subsequent trials would not have been conducted. The new treatment would have been accepted in much the same way that penicillin was accepted years before.\textsuperscript{10} Even standards of evidence that allowed the first partially randomized trial to count as significant evidence of efficacy would have eliminated the need for any further RCT and prevented the loss of many lives.\textsuperscript{11}

Several years after this controversy had blown over, lead researchers Robert Bartlett and Richard Cornell responded to an article by statistician Richard Royall, in which Royall calls the later ECMO trials unnecessary and unethical. Bartlett and Cornell acknowledge that they felt “compelled” by their colleagues to produce RCT evidence for ECMO and conclude their response to Royall by saying that in future research investigating the use of ECMO for pediatric and adult respiratory failure, “If we propose an approach without randomization to conventional therapy, we certainly will consider calling on Dr. Royall to argue our case with the NIH study section, the insurance carriers and editors of scientific journals.”\textsuperscript{12} In this statement, Bartlett and Cornell expose some of the pressures felt by medical researchers; when RCTs are the gold standard researchers feel compelled to produce RCTs – even when for a particular case the RCT method may be unnecessary, inappropriate, or even unethical.

The ECMO case forcefully demonstrates the impact of standards of medical evidence. This is not an exceptional case, though it is one of the most dramatic. Countless lessons can be gleaned from this case. Many of the more critical lessons require detailed explanation of the sort that appears later in this dissertation. For now I will draw attention to the impact of standards of evidence on the health and well-being of research subjects and patients generally. Standards of evidence had life and death implications for the neonates in the later ECMO trials. In light of these very serious consequences, the standards of evidence requiring RCTs had better be justified. And yet quite often these standards are not explicitly justified. Many of the same criteria for evidence identified by the researchers in the ECMO trials exist today. In fact, they are arguably more entrenched today than they were in the early days of clinical epidemiology when the ECMO case was occurring. Insistence on RCTs, the devaluation of pathophysiologic

\textsuperscript{10} Ligon. The pivotal experiment, conducted in 1939, was an animal trial on eight mice. Within the 24 hours the four mice assigned to the control group were dead of streptococcal infection, while the four penicillin-treated mice were alive. This was treated as “miraculous” evidence, and within a year expanded animal trials were written up and published in the highly regarded medical journal, \textit{The Lancet}.

\textsuperscript{11} According to Worrall, this saga continued: a third RCT was performed at the request of statistician Stuart Pocock. Like Worrall, I do not see the need to belabour this point so will not discuss this last trial in any detail.

\textsuperscript{12} Bartlett and Cornell, 65.
reasoning, and the disdain for historically controlled trials and experiential evidence – these are all themes found in evidence-based medicine. The lack of explicit justification for these standards is one of the central concerns of this dissertation.

The ethical implications of standards of evidence are overwhelming when considered in the context of the tens of thousands of people who are research subjects in a given year, but are positively frightening in light of the even greater number of people who depend on the results of research for their health and well-being. For every loss of life in the ECMO trials there were countless other losses in hospitals around the world while physicians waited for conclusive evidence of the efficacy of ECMO. For every unnecessary trial there are victims within the trial and in society generally. Research that addresses the wrong questions, or investigates good questions with inappropriate methods, will ultimately waste time and resources, risk lives, and may not be able to stand up to ethical scrutiny.

2. Standards of Evidence and the Rise of Evidence-based Medicine

As the ECMO case demonstrates, the question of what counts as good evidence in medicine is of critical importance to physicians, philosophers and, ultimately, all of us as potential patients. One attempt to determine standards of evidence for medicine has been made by proponents of the evidence-based medicine (EBM) movement, as outlined in the ground-breaking 1992 article, “Evidence-based Medicine: A new approach to teaching the practice of medicine,” and revised in the oft-cited clarificatory article four years later, “Evidence-based Medicine: What it is and what it isn’t.”13 According to the latter, EBM is “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients.”14 In spite of a deluge of critiques in a range of medical journals, the EBM movement has been highly successful in capturing the attention and allegiance of many physicians, journal editors, administrators, and funding agencies, and is now a pervasive presence within health care systems worldwide. At the heart of the movement is a hierarchical ranking of research methodologies – the “evidence hierarchy” – which was created to assist researchers in assessing the relative value of evidence produced by various research methodologies, and to assist physicians in assessing

the results of research so that they might use it in practice. This hierarchy has been left largely
intact throughout the various iterations of the movement’s ideology, and contains a great deal of
the actual content of EBM. The hierarchy of interest in this dissertation – that of medical
treatments – places meta-analyses of randomized controlled trials (RCTs) at the top as best
evidence. Located at the bottom of the hierarchy are the sources of evidence regarded as least
reliable, such as case studies and anecdotal evidence. In this dissertation I will provide a
description and critical analyses of the current standards of medical evidence represented by
EBM. I intend this analysis to extend beyond the current focus on EBM and address questions
about standards of evidence and their justification more generally. Thus while EBM is the focus
of my attention because of its prominent role in determining current standards of evidence, the
analysis will be relevant to the justification of any future standards of medical evidence.

3. Locating the Project within Philosophy

3.1. Role of Philosophers

As the EBM ideology is rapidly and rather uncritically adopted in medical settings around the
world, as well as integrated into new domains (“evidence-based practice” is now common in
public policy, social work and some domains of complementary and alternative medicine, for
instance), concern about the assumptions, implications and epistemological limitations of such
an approach are mounting. Brian Haynes, one of the leading proponents of EBM, has issued a
challenge:

One hopes that the attention of philosophers will be drawn to…the continuing debate
about whether EBM is a new paradigm and whether applied health care research findings
are more valid for reaching practical decisions about health care than basic
pathophysiological mechanisms and the unsystematic observations of practitioners.16

15 Ibid.
16 Haynes, “What kind of evidence is it that Evidence-Based Medicine advocates want health care providers and
consumers to pay attention to?” 3.
The demand for philosophical attention to EBM is not only coming from the medical field. In a paper on randomized controlled trials and research methods in medicine, philosopher John Worrall writes:

This is an area to which philosophers … have given relatively little attention. Yet there are few areas to which they might contribute that have such enormous practical significance – involving, as it does, issues that are often literally matters of life and death.\(^1\)

I share Worrall’s optimism regarding the value of philosophy and agree with Haynes and other EBM proponents when they suggest that philosophical perspective is just what is needed in current debates over standards of evidence in medicine. Contrary to popular images of armchair philosophers and secluded ivory towers, philosophical investigation can contribute in meaningful ways to many avenues of inquiry, including the evaluation of standards of evidence employed by various scientific communities. Traditionally, philosophical analyses of issues in the domain of medicine have been a part of the sub-discipline of bioethics.

### 3.2. Bioethics and Philosophy of Medicine

The field of bioethics first emerged in the late 1960s as a product of several different academic disciplines including philosophy, theology, medicine and law. It was charged with the weighty task of uncovering and evaluating the ethical controversies raised by medical practice and medical research. Because philosophy had for many centuries been the primary location of secular academic debate on metaethics (the delineation and analysis of moral concepts, reasoning, and justification), normative ethics (the determination of how moral agents should behave), and practical ethics (the application of normative ethics to practical problems), bioethics was from the outset seen primarily as a sub-discipline of philosophy. As it evolved into a mature discipline over the years, it retained the traditional philosophical emphasis on critical evaluation, conceptual analysis, and the careful development of ethical theories.\(^2\) While bioethicists have shifted their focus from normative ethics to applied ethics (and perhaps back again), the focus of the field of bioethics has, unsurprisingly, been on the ethical issues arising in the medical

---

1. Worrall, “Controlling Clinical Trials: Some Reactions to Peter Urbach’s Views on Randomization.”
2. Borry, Schotsman and Dierickx, 58.
context. Too often overlooked in these discussions are the *metaphysical* and *epistemological* issues that arise in the medical context. Commitment to a sort of naïve realism, expectations that medicine is unproblematically scientific, and assumptions that the methods and models of science have been conclusively established are rife within the medical and bioethical contexts. These and other metaphysical and epistemological assumptions underlie standards of medical research and practice, and are in need of greater philosophical scrutiny.

Medicine as a field has suffered as a result of disproportionate attention to ethical issues. The ECMO case demonstrates how inattention to epistemological issues in medicine can have negative ethical implications. Concerns by researchers in complementary and alternative medicine about the metaphysical assumptions underlying medical research (for instance, mechanistic assumptions) have been largely ignored in the medical literature and scholars with the philosophical training to address such challenges, and to provide a defence of such assumptions, have not yet been drawn into the debate. There are countless other such examples, many of which will be raised later in the dissertation. While medical ethics – or bioethics today – is ostensibly an all-encompassing term for the ways in which philosophy and medicine intersect, it has been dominated by ethical concerns. This needs to change if we are to come to a more comprehensive understanding of health, disease, and the practice of medicine. A change would be most accurately reflected in a shift to “philosophy of medicine”, though a shift in content within bioethics would be a good start. In retrospect it is surprising that philosophers accepted the idea that ethical issues could be discussed in isolation from metaphysical and epistemological issues. Perhaps a more charitable reading of the evolution of the emerging field of bioethics would suggest that philosophers and other scholars began with the most pressing and life-threatening issues (those properly in the domain of ethics) simply because they were so urgent. Whatever the reason, a return to broader philosophical issues in medicine is overdue.¹⁹

While ethical issues are raised in much of my discussion, and are thus an important part of related future projects, the content of this thesis is largely epistemological. My aim is to provide a comprehensive epistemological assessment of current standards of medical evidence. The relationship between theories, hypotheses and evidence and the role of biases and values in

---

¹⁹ As a matter of disciplinarity, this means that bioethics should be reformed and renamed philosophy of medicine with a broader and more inclusive research agenda.
science have been the subject of much discussion in epistemology. I draw upon important new developments in epistemology in order to set out a framework for analyzing EBM.

3.3. Epistemological Framework

It might be possible to critically analyze current standards of evidence in medicine without proposing a specific epistemological framework. A traditional critical analysis would point out inappropriate goals set by EBM, and the gaps between the goals of EBM and the means settled on to achieve those goals. I decided to do more than this for the following reasons. There have been hundreds of papers written on EBM in the medical literature over the past fifteen years. A quick scan through the references for this thesis provides some indication of this large, and growing, body of critical literature. The critical analyses provided in the medical literature have, however, been piecemeal and scattered – in part because of the limited word-count of journal articles as compared to books or dissertations, but also because of repetition of the same central points in most articles. This indicates the need for a more comprehensive critical analysis. The length of this dissertation provides me with an opportunity to provide this more comprehensive analysis.

Moreover, the critical literature has tended to be just that: critical. Few scholars have pursued the constructive project of setting out principles and processes to guide the formation of standards of evidence. In order to pursue a project that is both comprehensive and constructive, that is, in order to provide a normative account of how standards of evidence ought to be justified in the medical context, I need epistemological concepts such as evidence, objectivity and justification. A critical philosophical assessment of standards of evidence requires an explicit account of evidence. Because concepts like evidence have many different meanings according to a wide variety of particular epistemological theories, the selection of one theory provides me with a basic understanding of these concepts from which to build my own framework. The development of a particular epistemological theory allows me to provide a more detailed assessment of EBM.

It is my hope that the epistemological framework utilized in this dissertation provides the opportunity not only for critical evaluation, but for deep critical evaluation of EBM in a way that identifies the underlying epistemological assumptions of the movement and the changes that
would be necessary in order to improve current and future standards of evidence. To use a medical analogy, this is akin to the difference between diagnosing a syndrome – which is essentially a cluster of symptoms with a name (for instance, irritable bowel syndrome) – and diagnosing any disease with an established etiology (for instance, a streptococcal infection in the throat). Not only is the second diagnosis more accurate, since it identifies the root of the problem (a certain strain of bacteria), it is more productive in that it makes the choice of treatment a lot more straightforward (though never entirely formulaic). Physicians know what kills streptococcal bacteria and what does not and how serious the infection can be if left untreated, and thus are able to have a productive and effective conversation with the patient about the best possible treatment. When a syndrome is identified, the conversation with the patient can occur, but there is a lot more trial and error required in determining appropriate treatment. Knowing the source of the problem is valuable. Returning to the project of this dissertation, I am concerned that a critical analysis of EBM that fails to sort out the complex and difficult epistemological issues surrounding evidence and justification will amount to a diagnosis of a syndrome, rather than a disease. This is especially true given that I will argue that the EBM movement is an attempt to delineate criteria for justification in the medical context, and thus is an attempt to constrain what counts as knowledge in medicine. Continuing with the analogy, I believe it will be much more effective to treat the problems with current standards of evidence in medicine if a particular epistemological framework is in place. The specific reasons for the choice of the social epistemological framework will be outlined in Chapter One.

3.4. Context

It is worth briefly tracing the relationship between the project of this dissertation and other projects in the social sciences and humanities, if only to clarify what I will not be addressing. A historically-oriented approach might look at how standards of evidence have been defined and altered throughout the history of medicine. Something like this has been offered by Jeanne Daly in her recent book *Evidence-based Medicine and the Search for a Science of Clinical Care.* In contrast, a sociological perspective might focus on the ways in which professions shape, and are shaped by, current standards of evidence and the sorts of power relations and institutional
practices that are enhanced, upheld or reinforced by different standards. Stefan Timmermans and Marc Berg have pursued a project of this sort in their book *The Gold Standard: The Challenge of Evidence-based Medicine and Standardization in Health Care.*

Each of these projects adds much to our understanding of EBM. The critical analysis of this dissertation is complemented by such projects, and they have contributed greatly to my understanding of EBM as a historically-located and socially-driven movement. I believe, however, that our understanding of issues raised by standards of evidence in medicine would be incomplete without attention to the strictly epistemological issues raised by EBM. In this dissertation I do not focus on the historically descriptive question, ‘how did such standards come about?’ or any of the possible social questions, including, ‘what interests are upheld by current standards?’ Instead I begin by answering the descriptive question, what *are* current standards of evidence in medicine? I then provide a detailed critique of these standards. Finally, I address the more difficult normative question, how *should* we determine standards of evidence in medicine?

It has often been said that when the time is right for an idea, thoughtful people will converge on it. It is a hopeful sign, then, that scholars in the humanities have begun to address some of the issues raised in the medical context in recent years. Kenneth Goodman has written on *Ethics and Evidence-based Medicine* (2003) and there have been two other efforts by philosophers to offer critiques of the epistemology of EBM. Doctoral theses by Robyn Bluhm (University of Western Ontario, 2005) and Maya Goldenberg (Michigan State University, 2007) are two such projects.

While there are similarities between these two projects and the general aim of this dissertation, my work is unique in three important respects. First, by making use of a social epistemological framework I am able to frame my critique in a way that differs from both Bluhm and Goldenberg. Bluhm offers a more historically-oriented critical analysis, and Goldenberg offers a “Davidsonian” analysis that draws upon work by Sharyn Clough. While Bluhm’s historically-informed analysis does an excellent job of identifying the forces that have shaped EBM and Goldenberg’s “Davidsonian” account ties ethical issues more centrally into concerns about evidence, neither has the resources to provide specific suggestions on how to restructure social institutions and practices in order to improve standards of medical evidence.

---

21 Timmermans and Berg.
The social epistemological framework I have chosen allows me to provide just these sorts of detailed constructive suggestions. Second, I offer a thorough discussion of the standards of evidence employed in Complementary and Alternative Medicine (CAM) – a project that both Bluhm and Goldenberg have recognized as an important element in the critical constructive project, though which neither of them has addressed in detail. Third, Bluhm, Goldenberg and I each come to slightly different conclusions about EBM. Though we are all in general agreement that the principles of EBM are problematic, we are inclined to locate the problems in slightly different places. This collegial disagreement will, I am certain, be the source of future critical and collaborative projects.23

4. The Argument Ahead

In what follows I provide an overview of the arguments advanced in this dissertation.

4.1. Chapter One: Social Epistemology

Traditional epistemologists examine the necessary and sufficient conditions for individual knowledge. Social epistemologists expand this project to include investigations into the social dimension of knowledge. Based on the strengths of social epistemological theories in managing social values, justifying social practices already common in the sciences, and providing explanations of the epistemological status of testimony, authority and expertise, I defend social epistemology as the most accurate and comprehensive approach to epistemology available today. Further, I specify criteria for the selection of the best social epistemological theory. According to my criteria, the best theory: focuses on justification, preserves individualistic intuitions, provides mechanisms for managing social values, and offers a robust account of the sociality of knowledge. On the basis of these criteria, I narrow the options down to one: critical contextual empiricism (CCE).

23 This collaboration has already begun. Robyn Bluhm, Maya Goldenberg and I have recently presented together on two panels: the Canadian Society for History and Philosophy of Science Annual Meeting in Saskatoon (May, 2007) and the First Biennial Conference of the Society for Philosophy of Science in Practice in The Netherlands (August, 2007). We have also been invited to act as guest editors for a second special issue on EBM in the journal Perspectives in Biology and Medicine. (Robyn Bluhm and I edited the first such special issue in 2005.)
4.2. Chapter Two: Critical Contextual Empiricism

CCE, which was originally proposed by philosopher of science Helen Longino, has been one of the leading theories in social epistemology for the past decade. In the second chapter, I outline the core elements of this theory and extract a method for analyzing standards of evidence. I develop a version of CCE that enables me to critically evaluate EBM on a variety of fronts. CCE is critical because it stresses the role of intersubjective critical debate in the production of knowledge. It is contextual because it takes the cognitive goals and practical aims of community members seriously when determining standards of evidence. And it is empirical because it stresses the importance of empirical adequacy as a criterion of knowledge (particularly scientific knowledge). I defend CCE from objections raised by Solomon, Richardson, Kitcher and Goldman by showing how CCE engages with the ‘real world’ of science, captures what is distinctive about scientific evidence, makes proper use of liberal democratic principles and avoids dogmatism. In the remainder of the dissertation, I draw upon my version of CCE in order to identify and critically evaluate assumptions underlying standards of evidence in medicine and to suggest and evaluate possible alternatives.

4.3. Chapter Three: Evidence-based Medicine

Proponents of the evidence-based medicine movement have established current standards of medical evidence. I take great care in this dissertation to represent EBM as accurately as possible. Straw-man critiques of EBM are commonplace, and it is not my intention to provide another critique of this type nor am I interested in developing a polemic against EBM. I am more interested in understanding the philosophical implications – the strengths and limitations – of the EBM approach. This involves charitably reading the statements of EBM proponents and, where possible, identifying elements of EBM that are vital to the future of medicine and worth maintaining. Toward this end, I begin with a historical survey of the EBM movement, from its origins in clinical epidemiology at McMaster University in the early 1990s to its current status as the model for medical research and medical practice in health-care systems worldwide. I then

---

24 Longino, *Science as Social Knowledge*.
isolate and describe the central elements of the most recent version of EBM. Finally, I identify the core background assumptions of the EBM movement – what, at least ostensibly, makes EBM different from any other kind of medicine.

4.4. Chapter Four: Critiquing the Background Assumptions of Evidence-based Medicine: Part I

According to the version of CCE I defend in Chapter Two, the most insidious sources of bias in scientific research are background assumptions. In this chapter I critique two of the background assumptions of EBM: the assumption that medical practice should be “more scientific” and the assumption that evidence from clinical research (rather than authority or pathophysiology) should form the basis of medical decision-making. With respect to the first assumption, I argue that the characterization of science underlying the drive to make medical research more scientific is narrow and inaccurate. In addition, scientific approaches to medical practice have not yet been shown to provide better patient outcomes. Further, medical practice is committed to principles of shared decision-making and individualized care which are in tension with the demands of science as characterized. With respect to the second assumption, I argue that while clinical research evidence was meant to replace authority in medical decision-making it has become a new form of authority, and that pathophysiologic reasoning, and for that matter, a number of other factors, are as ‘basic’ to medical decision-making as research evidence.

4.5. Chapter Five: Critiquing the Background Assumptions of Evidence-based Medicine: Part II

The critique of background assumptions continues in Chapter Five, where I evaluate the assumption that a hierarchy of evidence is necessary to guide medical research and practice. EBM is distinguished from traditional approaches to medical decision-making by the central position accorded to the evidence hierarchy. In light of the variety of possible evidence hierarchies, the particular version offered by EBM needs to be justified. Advocates of EBM have not been forthcoming on this issue. In this chapter I reconstruct possible justifications for the current version of the hierarchy. According to my analysis, the evidence hierarchy defines best
evidence as that which is: precise, randomized, unbiased, quantified and relatively certain. I critically evaluate each of these characterizations of good evidence and consider whether they can provide justification for the evidence hierarchy. I argue not only that this particular hierarchy is unjustified, but further that any attempt to hierarchically rank research methods is incoherent since it amounts to an unjustified restriction of medical knowledge.


The remainder of the dissertation outlines structural changes that could be made to improve the standards of evidence used by medical researchers. According to my version of CCE, diverse perspectives are needed in order to ensure active critical debate in knowledge-productive communities. In the first section of this chapter, I argue that EBM contributes to a restriction on methodological and theoretical diversity within medicine. In order to demonstrate the value of diverse perspectives, and the importance of responsiveness to such perspectives in medicine, I outline recent developments in complementary and alternative medical research. While some of the research conducted by researchers in complementary and alternative medicine is of poor quality, and is rightly dismissed by the mainstream research community, this is not always the case. Using specific examples, I show how innovations proposed by alternative medical researchers could be of significant value to the mainstream medical community. These innovations include new, more patient-centered outcome measures and novel forms of the RCT. Furthermore, the insistence of complementary and alternative medical researchers on the value of underused but highly ranked research methods (such as the n-of-1) as well as lower-ranked research methods provides an indication of the way in which standards of evidence could be restructured non-hierarchically.

4.7. Chapter Seven: The Future of Evidence-based Medicine: Interactive Objectivity and the Open Science Movement

In the final chapter I take a closer look at the claims to objectivity made by EBM proponents. I draw upon a recent philosophical catalogue of objectivity proposed by Heather Douglas in order
to characterize the sense(s) of objectivity implicitly and explicitly assumed by proponents of EBM. I raise some concerns about the potential dangers of relying exclusively on these senses of objectivity. I argue that an over-reliance on *procedural* objectivity has led proponents of EBM to the false belief that methods alone (narrowly construed) can secure objectivity, and also to the related and even more problematic belief that guidelines produced on the basis of the evidence hierarchy provide an objective basis for medical decisions. Finally, I draw upon the CCE account of interactive objectivity in order to suggest ways in which standards of evidence might be improved. I provide a detailed account of interactive objectivity and outline the sorts of interventions in medical research and practice required by such an account, including: elimination of confidentiality agreements signed by researchers, disclosure of competing interests in publications, prohibition on ghost authorships, public registration of research, open peer review, open access to publications, open submission processes, increased public funding of research, less reliance on guidelines, greater awareness of the collective nature of medical decisions and, most importantly, greater analytic training for medical professionals.

5. **Aim of the Dissertation**

The overall aim of this dissertation is not to specify a rigid set of requirements for standards of medical evidence, or worse, to offer a formalized and over-simplified process for the evaluation of research methods. This would repeat the mistakes made by the proponents of EBM. Instead, I draw upon the resources of critical contextual empiricism to evaluate current standards of evidence and to propose means for improving not only the content of standards of medical evidence as they exist today, but the process of producing and defending such standards in the future. This project is advanced with an awareness of my position as an outsider to the medical community, and with the sincere hope that my arguments will be critiqued and ameliorated by medically-inclined philosophers and philosophically-inclined medical professionals in the future. While the EBM movement has brought about many welcome changes in medicine, including a greater focus on the need to provide explicit reasons for treatment decisions and a more critical attitude toward traditional forms of authority, it has important epistemological shortcomings. I argue that the standards of evidence proposed by EBM enable and even encourage false and

---

26 Douglas, 453.
potentially dangerous illusions of objectivity and certainty within medicine. The standards divert attention from alternative viable and valuable sources of evidence, thereby distorting medical research and practice as well as medical knowledge generally.
Chapter One

Social Epistemology

Traditional epistemologists examine the necessary and sufficient conditions for *individual* knowledge. Social epistemologists expand this project to include investigations into the *social* dimension of knowledge.¹ In this chapter I define social epistemology and offer reasons to support the expansion of the domain of epistemology to social issues. I then delineate and defend a set of criteria for a good social epistemology. A good social epistemological theory: focuses on justification, preserves individualistic intuitions, provides mechanisms for managing social values and offers a robust account of the sociality of knowledge. On the basis of these criteria, I narrow the options down to one: *critical contextual empiricism*. Finally, I clarify the relationship between feminist, social and naturalized epistemology. This chapter serves as a foundation for the detailed description and expansion of critical contextual empiricism offered in Chapter Two.

1. Individual and Social Epistemology

Epistemology is the study of knowledge: more specifically it is the study of justified true belief.² The starting point for traditional epistemology has been the Cartesian query, what can *I* know? Thus, the basis for most epistemological investigations has been the individual – and not just any individual, but a particular conception of a radically isolated, unattached and disembodied individual.³ Attention is generally focused on solving problems associated with individual beliefs, usually related to the fallibility of various psychological faculties. This focus has shifted, though this does not mean that epistemologists are no longer theorizing about individuals.

---

¹ Of course, there is no consensus on the precise scope of ‘social’ or ‘knowledge’. The scope of these concepts will be discussed in greater detail below.
² Contemporary analytic epistemologists regard this as provisionally established, though it is known to have certain shortcomings (the infamous Gettier problems). Supplemental criteria or a reconceptualization of truth and/or justification may be necessary in order to overcome such problems. Knowledge as justified true belief has also been disputed by other communities, notably postmodern theorists. I will not be discussing these foundational debates further in this dissertation as they are the subject of considerable attention elsewhere.
³ This characterization of trends in individualistic epistemology may not extend perfectly to ancient and medieval philosophy, since during those periods epistemology was not demarcated as a distinct field of philosophy in the way it is today. The epistemological projects that were pursued often addressed metaphysical and other problems concurrently, and tended to have varied characterizations of the individual subject.
Rather, the scope of epistemology has broadened as attention to the conditions for social knowledge increased over the past three decades. Social epistemology, then, is the study of the social conditions of knowledge and knowledge-production. I am sympathetic with scholars who describe the shift toward social epistemology as “long overdue.”

All social epistemologists share a commitment to studying the social dimensions of knowledge. They differ significantly, however, in their views about how this investigation should proceed. Roughly, the difference is as follows. Classical social epistemologists believe that social epistemology is an extension of individualistic epistemology, and that investigation into the social dimensions of knowledge can be a fruitful addition to traditional projects. Advocates aim to correct for the overly individualistic tendencies of traditional epistemology. They retain a commitment to truth and/or justification and so are able to engage with the traditional canon in epistemology. On the other hand, anti-classical social epistemologists see the new field as a replacement for the misguided (not merely incomplete) enterprise of traditional epistemology. Advocates of anti-classical social epistemology focus on the contextual differences that shape different communities of knowers; they tend to discount the need for discussions about truth and justification, preferring instead to focus on belief-forming practices.

2. Social Epistemology

2.1. Past and Present

Although the social dimensions of knowledge have occasionally merited attention from the philosophers of the traditional canon, the first systematic exploration of issues in social epistemology didn’t occur until the 1970s. The term social epistemology, first proposed by Margaret Egan, referred to “the study of knowledge in society”, with a focus on “the production, flow, integration, and consumption of all forms of communicated thought throughout the entire

---

5 This distinction is a bit crude, but serves as a basic division of the domain. It is crude because social constructivists and others in the anti-classical camp might still be characterized as offering accounts of truth and justification – it is just that the accounts they offer are highly relativistic and/or naturalized. Goldman’s distinction here probably has more to do with the proximity of accounts of truth and justification to traditional accounts of the concepts.
social fabric.” The term was adopted into the literature in the social sciences (particularly in science and technology studies) and has gained a certain degree of notoriety there in recent years. The versions of social epistemology advanced by social scientists have (unsurprisingly) tended to be descriptive – they have been more concerned with the relationship between power and knowledge in contemporary scientific communities than the ideal conditions for the production of social knowledge. In what follows, I will focus on the philosophical approaches to social epistemology. This burgeoning field contains both classical and anti-classical social epistemologists. The creation, in 2004, of *Episteme*, a journal dedicated to publishing research in analytic social epistemology, as well as several recent social epistemology conferences attended by highly regarded analytic epistemologists, seems to indicate a general acceptance of the field that was not present even a decade ago. More philosophers are recognizing the importance of normative philosophical work in an area traditionally dominated by sociologists and scholars in science and technology studies. Add to this the convergence between feminist and social epistemologists and the community is experiencing a significant amount of growth.

### 2.2. Reasons for Social Epistemology

An epistemological theory that neglects the social dimensions of knowledge is incomplete. It also fails to provide guidance on the ideal structure and organization of knowledge-productive communities. Many epistemologists have recognized the importance of examining the social aspects of knowledge-production. There are four reasons why greater attention to these issues is warranted. Social epistemology is: 1) helpful for understanding testimonial knowledge; 2) responsive to the practice of knowledge-production; 3) fruitful for understanding the status of expertise; and 4) able to solve otherwise intractable problems caused by the presence of social values in inquiry. I will provide support for each of these claims below.

First, we might care about the social elements of knowledge because we want to understand testimonial evidence. From the moment we begin to understand language, we learn a

---

7 Shera, 86.
8 Steve Fuller, professor of sociology at Warwick University and author of one of the first texts on social epistemology, has been widely criticized for taking the stand as an expert witness in support of intelligent design, [http://education.guardian.co.uk/higher/profile/story/0,,1698284,00.html](http://education.guardian.co.uk/higher/profile/story/0,,1698284,00.html)
9 For a survey of the views represented in contemporary social epistemology see Alvin Goldman’s *Knowledge in a Social World*, 69-79 and *Pathways to Knowledge: Private and Public*, 139-204.
great deal through the testimony of our caregivers. The question of whose testimony we should trust is thus an important one. Traditional (individualistic) epistemologists have for some time recognized the value of philosophical work on testimony (Hume, for instance, offered a compelling argument for distrusting testimony). This has been the thin edge of the social wedge in epistemology, so to speak. Issues of testimonial knowledge require that traditional epistemologists refocus their attention on the social character of some aspects of knowledge production. For instance, when should we believe the testimony of a medical professional or bioethicist? What about the testimony of a media reporter avowing the health benefits of a new drug? Or the testimony of patients in an online chat group? Given that a significant portion of our knowledge seems to come to us this way, this is an important question for epistemologists. Epistemological theorizing on testimony attends to the dynamics of interaction between and among individuals. This is sometimes classified as a weak form of social epistemology. It provides a starting point for further, more robust social investigations. The remaining three reasons address these more ambitious, and more interesting, projects.

Second, investigations into social features of knowledge-production allow epistemologists to understand and explain the function of social conventions already in use in paradigmatic knowledge-productive communities, such as scientific communities. Building on this, such investigations also allow epistemologists to modify these social features so as to improve knowledge production. Science is a good example of a social knowledge-producing enterprise. The education of scientists is social, scientific ideas and instruments are shared, and scientists work in networks and communities. Science requires by its nature the presence of two or more individuals. Certain structural and inherently social features of scientific communities enhance the production of knowledge. These include scholarly meetings and conferences, congresses, seminars, round-tables and workshops, as well as journals (peer-reviewed and otherwise), requirements that researchers perform literature reviews before beginning a project (in order to link work to related work that was done in the past), third-party attempts to replicate research, public lectures, forums for interdisciplinary work, and many other specialized mechanisms in particular areas of science (for instance, grand rounds in medicine). These are important elements of scientific inquiry: they provide an opportunity for critical debate and outside evaluation of research, as well as an opportunity to disseminate research within the community in order to trigger new ideas. Since these social conditions can be arranged in better
or worse ways, social epistemologists argue that attention to the social conditions of knowledge
cannot be ignored in the construction of a comprehensive epistemological theory. Social
epistemologies are more comprehensive than individualistic epistemologies because they
incorporate normative standards that guide the interaction of individuals as members of
communities.

Failure to attend to the social structures inherent in knowledge production can lead to
problems. Imagine a scientific community organized in a way that prevents critical dialogue
among community members. Such a community would likely limit the publication and
distribution of research, ensure secrecy by limiting the presentation of material at conferences
and other public events, and encourage individual scientists to pursue projects without significant
collaboration or assistance from others in their own or other disciplines. The number of journals,
conferences, public lectures and collaborative projects would be significantly less than what we
see in science today. We take for granted that the features of scientific communities we have
today are well suited to the production of knowledge. But on what grounds do we make this
claim? Unless we engage in discussion about why some structural features (such as critical
debate) are epistemically valuable, we will not be able to justify our choice of one of these
communities over the other. When forces outside of science, such as those representing powerful
economic interests, demand changes to scientific communities such as rules of confidentiality in
research, it is social epistemological arguments that justify resistance.\textsuperscript{10}

Third, the investigation of social factors in knowledge-production may be fruitful for
other problems we face in the scientific realm. The status of experience and expertise as reliable
grounds for knowledge is a persistent concern in many areas of science. In medicine, for
instance, the expertise of physicians was traditionally one of the most highly sought-after sources
of knowledge. In the medical literature today many of the critics of evidence-based medicine
(EBM) decry the loss of the art of clinical judgement and individual expertise. They argue that
EBM is driven by a desire to bypass individual expertise and experience, and ultimately
physician judgement, because such judgement is seen as idiosyncratic and prone to error.
Understanding what value should be placed on the expertise of individual physicians will require
attention to the ways in which expertise and experience are established within the medical
community, and will ultimately require the attention of social epistemologists.

\textsuperscript{10} I discuss these issues further in Chapter Seven.
Fourth, we need to pay attention to the social factors in science because we will then have the resources to resolve one of the big questions in science today: what role should social values play in scientific inquiry?\(^{11}\) Values are pervasive in scientific inquiry. Scientists need to understand what to do about this state of affairs. As extensive research in sociology and science and technology studies over the past half century indicates, social values are pervasive in science from macro-level political and economic forces to the micro-level decisions made by individual researchers.\(^{12}\) On the macro-scale, to use a popular example, massive government and commercial investment in scientific research following World War II (for instance, in the Manhattan project, the Apollo project, and the RADAR project) led to the era of “Big Science”. This changed the character of science: projects became much larger and the number of scientists working on particular teams expanded. As many historians have shown, this impacted not only the content of science (since research money was directed to projects likely to advance national and commercial interests) but the practice of science, as scientists no longer had independent control, or in some cases even understood, the data they produced.\(^{13}\) Similar social values, for instance those driving the recent Human Genome Project, are thought to influence science today.

On the micro-level, sociologists and anthropologists have demonstrated that individual scientists make dozens of ‘small’ decisions in the course of research. These seemingly insignificant decisions can have influence what aspects of the data are found “most interesting” and pursued, which are declared “successful,” and how the data are presented in publication (for instance, data are often presented in ways that downplay the active role of the researcher).\(^{14}\) Social epistemologists aim to offer accounts of knowledge-production that reflect our current understanding of the role of values in science without deferring to those who would suggest that, because values are ineliminable in inquiry, scientific investigations do not get at anything meaningful or real. Similarly, they refuse to defer to those who suggest that values can be isolated and extracted from scientific inquiry. Instead, social epistemologists exploit the sociality

\(^{11}\) I will explain the nature of social values in Chapter Two. For now it will suffice to note that values can be negative or positive.

\(^{12}\) See Brown’s *Who Rules in Science* for a discussion on the various ways in which values permeate science, and for an overview of the famous cases of the past two decades. According to Brown, values are beliefs that are not empirically testable. The term “values” is used broadly here to refer to both epistemic and non-epistemic values or, in Longino’s terminology, constitutive and contextual values. The primary concern has been with the negative influence of contextual (or social) values on science. This distinction is discussed further in the next chapter.

\(^{13}\) Shapin and Schaffer’s *Leviathan and the Air Pump* is an excellent investigation of the ways in which large scale institutional and professional interests influence science.

\(^{14}\) Knorr-Cetina.
of science in order to solve what are otherwise intractable problems. In particular, social epistemologists are able to find middle ground between the two extremes in the science wars and propose a constructive role for values in science. This is a significant achievement and provides support for further research into social epistemology.

3. Criteria for Choosing a Theory Within Social Epistemology

In the Introduction I explained my reasons for using a particular epistemological framework for the critical project of this dissertation. In order to proceed with this project as outlined, I must identify the best social epistemological theory. To begin, I advance a set of criteria for a good social epistemology, and explain and defend the criteria. My development and articulation of these criteria reflects my commitment to finding a social epistemology that is robustly social yet which recognizes the value of explaining our most basic intuitions about individual knowledge. These criteria also indicate my commitment to honouring promises to deal with the problem of social values in knowledge-production (as noted above in the justifications for social epistemology generally). Once I explain and defend the criteria, I use them in my assessment of the available social epistemological theories.

First, the theory must explicitly address issues of justification. Jaegwon Kim explains the focus on justification in epistemology by noting that,

Neither belief nor truth is a specifically epistemic notion: belief is a psychological concept and truth a semantical-metaphysical one. These concepts may have an implicit epistemological dimension, but if they do, it is likely to be through their involvement with essentially normative epistemic notions like justification, evidence and rationality.

In order to justify our beliefs we must be able to provide reasons for them. We must offer good reasons to think that our beliefs are true (or meaningfully true, or conform to reality, or whatever pluralistic goal we have set). Kim goes on to explain how it is that the element of justification brings normativity to epistemology,

---

15 A more detailed examination of these issues is offered in Chapter Two.
16 Kim, 383.
Justification is what makes knowledge itself a normative concept... If a belief is justified for us, then it is permissible and reasonable, from an epistemic point of view, for us to hold it and it would be epistemically irresponsible to hold beliefs that contradict it. If we consider believing or accepting a proposition to be an ‘action’ in an appropriate sense, belief justification would then be a special case of justification of action, which in its broadest terms is the central concern of normative ethics.\(^\text{17}\)

My interest is in normative social epistemological theories. Because this project is a critical evaluation of current standards of medical evidence, and because I aim to provide constructive suggestions on the improvement of these standards, it is important that the social epistemological theory I draw upon be normative. As mentioned in the introduction, this is not a social scientific project, and my aim is only descriptive insofar as accurate description contributes to a more accurate critical analysis. As Kim points out, a focus on justification is necessary to bring out the normative dimension of an epistemological project. As a result, I will be interested in social epistemological theories that offer a detailed account of epistemic justification.

The second criterion is that the social epistemological theory preserves our individualistic intuitions. In other words, I am interested in a classical social epistemological theory. It is possible to recognize the importance of the social features of knowledge-production without denying that individuals can have knowledge and throwing out hundreds of years of traditional individualistic epistemology. I trust that it will not take too much effort to convince the reader that individuals can have knowledge since that has been the starting assumption of epistemology for thousands of years. While the inclusion of social features in our accounts of knowledge expands the project of epistemology into new domains, this should not mean the wholesale rejection of all important advances in traditional epistemology. The attribution of primary knowledge to individuals is a critical component of a good social epistemology. I am committed to retaining this component. Holistic accounts of communities as “super-entities,” or as the only subjects of knowledge, will fail this criterion.\(^\text{18}\)

Third, the social epistemological theory must provide mechanisms for managing social values. It should be able to tell us whether, for instance, secrecy is a good social value when it comes to producing knowledge. One of the strengths of social epistemology outlined earlier is

\(^\text{17}\) Ibid.  
\(^\text{18}\) Some social epistemological theories attribute belief-states to communities. Because beliefs have traditionally been attributed to only to individuals (with minds), the claim that communities have beliefs seems to imply that there is some super-entity, or collective being, to which beliefs are being attributed.
that it is an area of epistemology well positioned to offer solutions to the question of what to do with social values in inquiry. A theory that failed to do so would weaken the support offered for pursuing social epistemology in the first place. Because most social epistemologists do at least attempt to solve this problem, I am particularly interested in theories that offer detailed, practical mechanisms, where those mechanisms are well-defended.

Fourth, and last, the theory should offer a robust social account of justification. That is, the theory should have a comprehensive account of the relationship between the social and the rational. This requirement acts as a counterpoint to the second criterion: while the second criterion ensures a theory is not too social this criterion ensures that it is, still, robustly social. Epistemologists have traditionally been interested in the psychological faculties thought to lead to reliable beliefs. While desires, emotions, prejudices and imagination have been critiqued as unreliable bases for belief-formation, the traditional list of good sources includes: introspection, memory, perception, reason and (sometimes) testimony. So, for instance, beliefs based on perception are encouraged since they more reliably lead to knowledge than beliefs based on imagination. Social epistemologists suggest that we need not limit ourselves to psychological (i.e. individualistic) factors when we identify the reliable faculties that are available to us. In fact, this is already partially recognized by the inclusion of testimony as a reliable source upon which to form beliefs about the world.

If we examine the knowledge-productive processes common in the sciences we can see the possibility for further socializing our understanding of both perception (observation) and reason (and maybe, with recent developments in the cognitive sciences, memory). Philosophers of science usually describe perception (observation) and reasoning as the core elements of scientific practice. That is to say that observation and reasoning are generally accepted to be knowledge-productive practices. Deductive-nomological models of scientific explanation, for instance, draw on assumptions about the ideal relationship between the products of reason (theories) and empirical observations. In practice, attention to the sociality of knowledge will

---

19 Traditionally, social mechanisms were thought of as non-rational or even irrational. A theory that provides an understanding of the rationality of certain social processes moves beyond this traditional dichotomy. Such a theory will also suggest ways in which faculties traditionally thought to be free of social influence are actually dependent on social factors. See Longino, *The Fate of Knowledge*, for a detailed discussion of the relationship between the rational and the social.

20 Steup.

21 The socialization of observation and reasoning are discussed below. Work in the cognitive sciences and philosophy of mind on “external memory” seems promising as well. See for instance Olick and Robbins.
require that the account not be limited to explanations of testimony. While testimonial knowledge is a significant ‘gateway’ topic into social epistemology, theories that stop there fail to take the challenge of sociality seriously. Robust understandings of the sociality of observation and reasoning are a critical component of a comprehensive social epistemology.²²

4. Applying the Criteria

I will now use these criteria to narrow down the available options with which we are presented. In doing so I offer a cursory evaluation of some of the most prominent social epistemological theories, including those offered by Goldman, Anderson, (Miranda) Fricker, Latour, Kukla, Fuller, Nelson, Solomon, Cody, (Elizabeth) Fricker, Burge, Foley, Lackey and Longino.²³ I argue that Longino’s account best meets all four criteria.

4.1. Focus on Epistemic Justification

Anti-classical social epistemologies fail my first criterion because, by definition, they fail to engage with questions of epistemic justification. This means that theories by Bruno Latour, Andre Kukla, and Steve Fuller, among others, are set aside. Advocates of classical approaches to social epistemology tend to divide down two lines: those interested in the social conditions of truth, and those interested in the social conditions of justification. Most social epistemologists focus solely on one or the other of these concepts. As outlined above, social epistemological approaches that incorporate social accounts of justification will be of particular interest in this dissertation. This does not preclude the eventual consistency of such accounts with truth-oriented social epistemologies (such as Alvin Goldman’s ‘veritism’, and veritistic theories offered by Elizabeth Anderson and Miranda Fricker).²⁴ Bringing the two elements of classical social epistemology together is an important project, but not one developed here.

²² As we will see in the next chapter, it is this robust account of the sociality of observation and reasoning that will provide the grounds for solving the ‘problem’ of values in science.
²³ I must stress the cursory nature of these evaluations. It would take several dissertations to address all of these theories in detail, and the focus of this dissertation is elsewhere.
²⁴ Goldman’s veritistic theory of social epistemology has now been fully developed in the well-regarded treatise Knowledge in a Social World in 1999. Goldman is also the editor of the new journal Episteme: A Journal of Social Epistemology, which publishes articles in classical social epistemology.
4.2. Preservation of Individualistic Intuitions

The second criterion is that the theory must preserve our individualistic intuitions.\textsuperscript{25} It must permit the attribution of knowledge to individuals, though it may socialize the process leading to knowledge. In order to understand the position of social epistemologists on individual knowledge, it is first important to ask, what is distinctly social about social epistemology? Responses to this question can be grouped into two categories: those that argue for the existence of collective entities and those that argue for greater attention to inter-individual relationships.\textsuperscript{26} I will outline the two most common positions on the nature of the social in what follows.

4.2.1. Collective Entities

Advocates of the broadest definition of the social suggest that collective entities such as communities, groups and teams are capable of contributing to inquiry as a unit (not merely as individuals together, or as individuals-in-communities, but as a new entity). Thus we speak of Catholic beliefs, the doubts or convictions of juries, the actions of corporations, and so on.\textsuperscript{27} There is significant debate about the details of this understanding of the social including, among other things, whether all or merely some knowledge is a collective product. This understanding does seem, at the very least, to require the attribution of belief-like qualities to new collective entities or super-entities. But this raises a number of challenges, one of which is: how can we

\textsuperscript{25} Anti-classical social epistemologist will also fail this second criterion.
\textsuperscript{26} Very weak accounts of the social, which focus only on the role of social interests, are not the focus here. In such accounts, social interests have tended to be contrasted with rational interests, and so attention to social elements of knowledge in philosophy has been largely restricted to accounts of problematic social interests (biases) and ways of eliminating such interests. The idea is that we need to pay attention to social interests only insofar as we need to create mechanisms for eliminating them from inquiry. The focus of such mechanisms is on controlling the interests of individuals. This (weak) sense of ‘social’ in epistemology is really just an aspect of traditional epistemology, with some concern about corrupting social influences on individuals and individual mechanisms for limiting these corruptions. This understanding of ‘social’ allows individuals to produce knowledge regardless of their relationship to any other individuals, and independent of their membership in any particular community, as long as they have devised mechanisms for limiting the influence of social interests on their reasoning. As such it remains very individualistic.

\textsuperscript{27} Goldman uses a quote by Sandy Berger, former National Security Advisor in the United States Government, to make this point: “We’ve learned since 9/11 that not only did we not know what we didn’t know, but the F.B.I. didn’t know what it did know.” The F.B.I, as an integrated collective entity, didn’t know what many of its individual members knew (i.e. that terrorists were engaging in flight training in the United States). In this example, the collective entity is attributed beliefs that are not only distinct from those of its constituent members, but actually in opposition to many of those beliefs. Goldman, “Social Epistemology.”
attribute beliefs to collectives when beliefs have traditionally only been attributed to individuals? Determining the appropriate ontological status of a collective mind is particularly problematic. I do not have the space to address the details of this problem here. Instead, I will provide a quick example of the sort of account that runs into this problem.

Lynn Hankinson-Nelson’s holistic feminist empiricism draws upon Quine’s web of knowledge to suggest that evidence achieves its status from its relationship to other elements of the web of knowledge (other values, hypotheses, theories, reasons, and empirical observations). Knowledge is produced by the community and the community, as a sort of holistic entity, is what can be said to know. Individual knowledge only follows as a consequence of community knowledge (individuals have derivative knowledge only). This account has a very strong sense of the social – going so far as to suggest that it is only communities that have primary knowledge. This erases, or at least seriously limits, the individual subject, and marks a significant departure from traditional epistemology. This means that Nelson’s social epistemology is too social by the criteria outlined above.

4.2.2. Inter-individual Relationships

In contrast to this holistic account, more conservative accounts of the social attend to the interactions and relationships among individuals. Advocates of attending to these inter-individual relationships argue that we need not focus only at the individual level or the community level. In between these two are the interactions that occur among individuals in communities. It is possible to develop rules, principles and social structures to govern these interactions. As a result, individuals are credited with knowledge (there is no need to posit any further entities), but they only achieve knowledge as a result of relationships they have with other individuals. While this position is consistent with the claim that individuals can have knowledge, it does not allow individuals to be described as self-sufficient producers of knowledge. Individuals know, but they only know as members of communities. Social epistemologists of this type specify conditions for the organization of knowledge-productive communities. Individuals in a particular community produce knowledge to the extent that their community meets the normative criteria. A good social epistemology will conceive of the social in the manner just described. As

---

28 Heidi Grasswick develops this idea in her essay, “Individuals-in-Communities.”
mentioned above, social epistemological theories that posit collective entities fail the second criterion.

4.3. Mechanisms for Managing Social Values

A successful social epistemological theory will provide mechanisms for addressing the role of social values in inquiry. Miriam Solomon’s *social empiricism* attempts to provide such mechanisms. Solomon’s theory is social because scientific rationality is “socially emergent” and empirical because it is focused on empirical success.\(^{29}\) In keeping with the naturalist tradition, Solomon proposes to examine the production of empirically successful theories in science using empirical methods. Solomon suggests that “decision-vectors” (all factors which may influence a decision, whether positively or negatively) must be appropriately distributed within a community.\(^{30}\) She divides influences on science into empirical decision-vectors (those that are connected to empirical success, for example, salience and novelty of predictions) and non-empirical decision-vectors (for example, ideology and peer pressure) and assigns weights to each. Scientific inquiry produces knowledge when the empirical decision-vectors are distributed equitably and non-empirical decision-vectors are distributed equally. Her account is distinctive in that it proposes a modified quantitative approach for balancing empirical and non-empirical factors in scientific inquiry. Her account also has a central role for rationality. Solomon argues for the importance of empirical success as a criterion of rationality, since empirical success is thought to be an indicator of truth. Communities are rational when they use empirical success as a measure for adopting or rejecting theories. Individuals are offered some credit on this account (preserving some of our intuition that individuals can have knowledge), in that they can hold beliefs that differ from those of the communities in which they are situated. Rationality is attributed to the community and to individuals, and the two can come apart. By the criteria outlined so far, Solomon’s account might seem promising.

Unfortunately, Solomon’s account encounters serious problems when it comes to the details. In particular, it has been extremely difficult to identify appropriate mechanisms for distributing decision-vectors – particularly while remaining naturalistic. Solomon’s desire to

---

\(^{29}\) Solomon, *Social Empiricism*, 12, 120.

\(^{30}\) Ibid., 53.
remain a naturalist is complicated by a number of clearly normative claims (such as principle of equity and equality underlying the distribution of decision-vectors). Unfortunately, the desire to preserve a naturalistic framework while making normative rules forces her to make some elaborate and unpersuasive claims that, as Psillos puts it, “[verge] on incoherence.” There are also concerns about the attribution of rationality at the community level (similar to the concerns raised about Nelson’s view). In addition, criteria for identifying decision vectors are regrettably absent from her account. As a result, her assertions regarding the number of decision-vectors in a particular historical case seem arbitrary. Finally, one wonders how necessary (and how practical) the quantified approach to balancing empirical and non-empirical factors really is.

Social epistemologists should provide mechanisms for managing social values, and Solomon attempts to do so. Her choice of mechanism, however, raises more questions than it answers. How are we to properly assign value to empirical and non-empirical factors? What happens if we purposely weight non-empirical factors differently in order to achieve outcomes we desire? Does it make sense to weight all empirical factors equally – are accuracy and fruitfulness always equally valuable? Does it make sense to weight all non-empirical factors equally – is political ideology comparable to religious devotion? Finally, how robust is the distinction between empirical and non-empirical factors? It has been the contention of many science studies scholars that the distinction between so-called epistemic and non-epistemic factors cannot be upheld. Similar concerns seem likely regarding empirical and non-empirical factors. In light of the challenges faced by Solomon’s account, particularly the lack of practical and defensible mechanisms for managing social values, this social epistemological theory fails my third criterion.

4.4. Robust Relationship between the Rational and the Social

I am looking for an account of knowledge that preserves our individualistic intuitions while ensuring a robust account of the relationship between the social and the rational. This is a delicate balance. There are a number of classical social epistemologists working on the justifications for testimonial evidence, including C.A.J. Coady, Elizabeth Fricker, Tyler Burge, Richard Foley and Jennifer Lackey. What is surprising is that this focus on the justifications

---

31 Psillos, 546. Another (largely negative) review is provided by Klee.
offered for testimony is not extended to other information-gathering faculties such as observation and reasoning. I would classify these accounts of justification as only weakly social because they neglect the potential sociality of these other faculties.\footnote{It may be the case that while these authors have focused on testimony they intend to expand the scope of their claims to other social features. In that case, I would be most interested to read what they have to say. My position here is that they do not go far enough, not that what they have done is wrong. It may turn out to be the case that these scholars develop excellent, more comprehensive, accounts in the future.} Robust theories will have to consider a deeper sociality. Explicit discussion of the sociality of observation and reasoning is required. Among the accounts of justification offered by classical social epistemologists, the prominent theories with what I will call a robust commitment to the social are: critical contextual empiricism (Helen Longino), social empiricism (Miriam Solomon), and feminist empiricism (Lynn Hankinson Nelson). These prominent epistemological theories have an integrationist project with respect to the rational and the social.\footnote{That is, they attempt to integrate the two.} Of these, we have already rejected Nelson and Solomon (above) based on other criteria. Helen Longino is interested in whether observation and reasoning might be socialized, and whether this socialization strips them of their justificatory power (she argues it does not). She is not suggesting that observation and reasoning are wholly social, but rather that they have a social dimension. In doing so, she elevates her theory above those focused on testimony and offers a more robust social epistemology. In what follows I will outline how Longino provides this more robust sociality. She does so by socializing observation and reasoning.

4.4.1. Observation

There are at least two ways in which observation is social, according to Longino. First, observation is not simple sense perception, but “an organized sensory encounter that registers what is perceived in relation to categories, concepts, and classes that are socially produced. Both ordering and organization are (dependent on) social processes.”\footnote{Longino, The Fate of Knowledge, 100.} The theory-ladenness of observation (that is, of concepts and apparatus) makes observation at least partly constituted by its theoretical context. This is a familiar claim in philosophy of science though Longino identifies an unfamiliar consequence of accepting observation as theory-laden: the socialization of cognition. Insofar as the community of knowers produces the concepts and apparatus.
necessary to observation, observation is at least partly social. Decisions about which aspect of an observed phenomenon is important, or what aspect should be observed at all, as well as progressive debate over the proper interpretation of an observation, all occur among scientists. What counts as an observation, and then what counts as an observation worthy of attention, are determined by the community in which the observation is made.

Second, observations must be stable across a variety of observers. Scientists are familiar with the requirement that their experiments are reproducible by others. This criterion of reproducibility, or repeatability, is grounded in assumptions about the fallibility of the observations of individual scientists (in particular contexts, at particular times) and thus the importance of outside access to research. Such a requirement aids scientists in determining that the observed regularity was not fraudulently created or unique to a particular setting, individual, or context. It is also meant to distinguish private knowledge that is not accessible to others (for instance, divine revelation) from public knowledge (that which is accessible to any person under the appropriate conditions). Science is the investigation of publicly accessible empirical phenomena. Scientific experiments are publicly accessible, and requirements of reproducibility provide assurance that the phenomena is public and that scientists are observing the same phenomena.\(^{35}\) In sum, the claim that observation is social amounts to a claim that,

\[T]\he status of the scientist’s perceptual activity as observation depends on her relations with others, in particular her openness to their challenge to and correction of her reports…\[T]\here is no way but the interaction of multiple perspectives to ascertain the observational status of individual perceptions.\(^{36}\)

Longino is not suggesting that individuals cannot engage in experiments as individuals, rather that perceptions will only acquire the status of observations in light of particular conditions.

\subsection*{4.4.2. \textit{Reason}}

\(^{35}\) A concern with this second claim is that scientists rarely bother to actually reproduce the experiments of others. This is evident in medicine, where large-scale RCTs are rarely repeated once they have been performed. In practice, it is often only the controversial results that are re-tested. For instance, research into alternative medical treatments is quite often subject to ‘re-confirmation’ by different groups of investigators (this is particularly common with homeopathy). This problem is usually addressed by claiming that it is the ‘in principle’ repeatability (not the actual repeated observations) that matters.

\(^{36}\) Ibid., 103.
The claim that *reason* is social gets to the core of Longino’s position. She focuses on justificatory reasoning as opposed to creative or constructive reasoning. This is important because there is a difference between the process of creative reasoning, which I think Longino would allow may often be entirely individual in nature, and justificatory reasoning, which is not. When Longino talks about justificatory reasoning, she means to include not only stereotypical reasoning-as-thinking, but the act of providing justificatory reasons for one’s beliefs. Thus it is not thinking that is socialized (it is difficult to imagine what would that look like) but the act of providing justificatory reasons.\(^{37}\)

Drawing on work by Steward Cohen and David Annis, Longino argues that justification is the practice of responding to objections. We see the rudiments of this position in the advice we give to undergraduate philosophy students – “once you’ve put forward your own position, be sure to consider an objection and reply to it.” No philosophy essay would be complete without this hypothetical interrogation by a critical outsider. Whether the justification is successful or not will depend on the community standards adopted in a particular context. If social rules governing the appropriate forms of criticism play a role in this account of justification, it appears as though justification is partly social. As Longino puts it, “the critical dimension of cognition is a social dimension” and further, “socializing cognition is not a corruption or displacement of the rational but a vehicle of its performance.”\(^{38}\) Again, this does not deny the role of *individuals* as knowers. Longino is suggesting that while individuals form beliefs through individual reasoning processes they only produce knowledge when that belief is justified – and justificatory reasoning is social. Individuals are said to have knowledge when their beliefs have survived critical interrogation by others in their cognitive communities. In this sense, though individuals reason to conclusions, the products of that justificatory reasoning are only honoured with the title of knowledge when external scrutiny has occurred.\(^{39}\) Of course we can all think, imagine and reason in our own

\(^{37}\) As Ian Hacking puts it: “reasoning is not, as I understand it, a purely sedentary art. It includes a lot of doing, not just arguing or thinking.” Longino would add that the activity of reasoning is undertaken by communities of individuals and is thus not only active but interactive. Hacking, “Statistical Language, Statistical Truth, and Statistical Reason,” 138.

\(^{38}\) Longino, *The Fate of Knowledge*, 106-7.

\(^{39}\) There is a question here about the appropriate endpoint of inquiry. In line with Peirce, I believe we can never know we have reached knowledge because we can never be sure we have subjected our beliefs to sufficient critical interrogation.
minds, but we cannot call that knowledge until it has been publicly justified under the conditions she sets out. Longino uses a nice example to make this point. She says,

*Of course,* Galileo and Newton and Darwin and Einstein were individuals of extraordinary intellect, but what made their brilliant ideas knowledge were the processes of critical reception. Comparing the fate of their ideas with those of a thinker of arguably equal intellectual power, Freud, demonstrates this.

Social factors are partly constitutive of, and serve to stabilize, knowledge. The socialization of cognitive practices such as observation and reasoning forces us to recognize the interdependence of individual subjects. Because normative rules can still be exercised on individuals and on communities, we are left not with empty relativism, but rather with a more robust account of the types of rules and practices that will have to be specified by a particular community in order for it to produce knowledge.

5. **Critical Contextual Empiricism**

Helen Longino offers an account of epistemic justification. She argues that justification always draws upon assumptions, and in order to ensure that assumptions are not arbitrary, we need to ensure that conditions for critical evaluation of assumptions are in place. These conditions are social – communities have to be sufficiently diverse, for instance. Longino’s account yields individual knowledge but insists that it must be produced intersubjectively (interactively with other individuals) to count as knowledge. Thus the community is important in the construction of knowledge but at all times this is grounded in the interactions between individuals in relationships with each other. There is no need to posit super-entities in order to have a novel and fruitful account of social factors in knowledge production. At the same time, Longino offers an account of the role of social factors in knowledge-production that incorporates sociality into rationality. This highly integrationist project has a robust role for the social, and so does not fail to be a strong social epistemology. I will explain her approach, respond to criticisms, and identify the key features that will be helpful in my evaluation of EBM in the next chapter. Before I do this, I will outline the relationship between Longino’s critical contextual empiricism (CCE)

---

40 Details of these conditions are provided in Chapter Two.
41 Ibid., 122. Emphasis in original.
and naturalized and feminist epistemologies, since Longino’s account is often classified under these other categories and because an understanding of how her account is situated will be helpful in the next chapter.

5.1. Naturalized Epistemology

CCE builds on several trends in traditional epistemology in the last half century, notably fallibilism, anti-foundationalism, pragmatism and naturalism. Of these, naturalism (in the form of reliabilism) is arguably the dominant philosophical position adopted by analytic epistemologists today. Naturalists hold a variety of positions, but are minimally committed to the integration of psychology into epistemology and the rejection of a priori knowledge. Why, then, isn’t Longino just a naturalized epistemologist? CCE is certainly influenced by naturalism to the extent that it takes seriously the workings of real scientists. But Longino is wary about a whole-hearted embrace of naturalism for a couple of reasons. First, naturalists tend to recreate the same individualistic accounts of epistemology traditionally adopted by the philosophers of the canon. Naturalized epistemologies draw on work in cognitive science and psychology in order to understand how individuals come to have knowledge. Longino wants more than this. She wants an epistemology that pays attention to the social factors influencing knowledge-production. This would require an expansion of the naturalist’s project to include empirical work by, for instance, sociologists and anthropologists. This is not a conceptual impossibility, though it does go against the general tide.

Second, and more pivotally, Longino is concerned that the naturalized epistemologies will fail to have a strong normative dimension. And, fundamentally, feminist social epistemologists such as Longino are not just concerned with identifying practices of oppression; they are interested in changing those practices. This strong political goal influences the choice of epistemological framework. As Longino puts it, naturalism in epistemology “presupposes that we know what we think we know and asks how.” But it is not at all clear to feminists that we know what we think we know. In fact, the work of science studies scholars indicates that, at least some of the time, we are misguided about what we think we know. Theories such as Longino’s

---

critical contextual empiricism are thought to strike a delicate balance between an overly-descriptive (thus conservative) and possibly individualistic naturalism, and the needs of feminist epistemologists as feminists concerned with emancipation and as social epistemologists concerned with interactions among agents. Of course, because the naturalistic landscape is more nuanced than ever before, it may be possible for feminist epistemologists to retain enough of the normative dimensions of science while remaining weakly naturalistic. The details of this negotiation are an interesting area for further inquiry, though I will not address them here.

5.2. Feminist Epistemology

Feminist epistemologies fail to fit the brutal caricatures often drawn of them by critics. Contemporary feminist epistemologies do not: essentialize women’s intuitive and emotional ways of knowing, hold that a particular privileged viewpoint is held only by women, retreat uncritically into relativism, or support the wholesale rejection of rationality. Elizabeth Anderson’s fantastic review “How not to criticize feminist epistemology” has gone a long way to (finally) ending the straw-man arguments and beating of dead horses that passed for critiques of feminist epistemologies in the past.\textsuperscript{45} The traditional division of feminist epistemology into three camps – empiricism, standpoint theory and postmodernism – was originally proposed by Sandra Harding in 1986. Twenty-odd years later the distinctions are not as distinct, nor as helpful, as they once were. In fact, there has been a great deal of convergence in feminist epistemology, with most theorists agreeing on core commitments to pluralism and rejection of the possibility of a “view from nowhere.” Some differences remain, but many of these are merely products of the different tools used in the different areas. Empiricists make use of analytic philosophy of science, standpoint theorists draw upon cultural theory, and postmodernists rely on literary theory. In addition, recent work by feminist pragmatists has added the tools of pragmatism to this landscape. Standpoint theory has met with significant challenges in escaping the circularity of its position and in overcoming the bias paradox and feminist postmodernism fails to advance normative prescriptions for the improvement of science. There has been a convergence on empiricist approaches in recent years.

\textsuperscript{45} Anderson, “Review”
Feminist empiricist epistemology is best described as a subset of social epistemology. Gender is a social category with similarities to other social categories (such as race and class). A feminist epistemology that reflects contemporary feminist understandings of the links between different forms of oppression will be anxious to expand discussion to all social factors in accounts of knowledge. The relationship between feminist and social epistemologists has been a productive and interactive one, with many feminist epistemologists identifying as social epistemologists and vice versa. The feminist character of much social epistemology is evident in the critique of oppressive epistemological practices (through case studies and sociological investigations) – often related to gender. Feminists have traditionally valued interconnectedness over isolation, both politically and epistemologically, and this general tendency has also been well integrated with the project of social epistemology.\textsuperscript{46}

Longino describes what it means to be a feminist empiricist epistemologist. I will quote her passage in full,

\textit{To do epistemology as a feminist is to engage the questions of epistemology with an awareness of the ways in which participation in socially-sanctioned knowledge production has been circumscribed, of the ways in which epistemological concepts like rationality and objectivity have been defined using notions of masculinity (and vice-versa), of the ways women have been derided as knowers, and of the need for alternative theoretical approaches to satisfy feminist cognitive goals. It is to ask how epistemology has participated in or sanctioned these disbursements of privilege and opprobrium and to ask whether the efforts to exclude women from knowledge generating activity has not also resulted in the exclusion from the analysis of knowledge of traits and capacities assigned to women (a shrinking of the conception of knowledge.) What is important for the feminine or the female here is the perspective it affords on the construction of the concept of knowledge and the window it opens on alternatives. But it functions as an object of reflection, not as a subject position.}\textsuperscript{47}

To paraphrase, Longino suggests that the project is not to create a specifically feminist epistemology but rather to do epistemology as a feminist – to think about the production of knowledge with an eye to the ways in which epistemology may be related to practices of oppression. This approach acknowledges the ways in which other forms of oppression (racism and classism, for instance) also influence knowledge-production. The key feature of the account

\textsuperscript{46} Grasswick, Webb, “Feminist Epistemology as Social Epistemology,” 190.

\textsuperscript{47} Longino, “In Search of Feminist Epistemology,” 475.
is awareness – of any and all potentially harmful distortions of our understanding of knowers, knowledge-production, and knowledge.

6. Summary

Helen Longino’s critical contextual empiricism offers an account of epistemic justification that preserves individualistic intuitions while offering a robust account of the sociality of knowledge. CCE does not, unlike other more radical epistemological accounts, suggest that everything we’ve learned in epistemology is useless because it is too de-contextualized, too objectified, and too individualistic. At the same time, the account is not shy in its critique of the ways in which traditional accounts neglected the social realm. But Longino does not just claim that individuals can justify beliefs to other individuals (as a limited focus on testimony might suggest); rather, she claims that justification just is a social process. In the next chapter I outline the key elements of CCE and defend it against objections. I also propose new interpretations of key elements of CCE. Once these arguments have been established, I proceed to my critical analysis of evidence-based medicine.
Two significant challenges to the traditional account of scientific inquiry arose in the twentieth century. The first was the argument that observations are theory-laden. The second was the argument that theories are underdetermined by data. Each of these claims was contrary to a traditional account of scientific methodology in which observations and reasoning were, among other things, free from the influence of subjective values. The last half-century of scholarship in philosophy of science and related fields has been marked by an attempt to come to grips with the problem of inescapable subjectivity. Since the scientific method, as our most reliable producer of knowledge, depends on its perceived impartiality for credibility, these two problems remain central to contemporary philosophy of science. In fact, the “Science Wars” of the past decade have been largely a battle over the appropriate place of values in scientific inquiry.¹

In *Science as Social Knowledge*, Helen Longino develops an account of scientific inquiry that attempts to save science from the relativism of social constructivism while acknowledging the inescapability of subjective values in science. Longino suggests that it is possible to acknowledge the value-laden nature of all human inquiry without abandoning the possibility of objectivity. Scientific reasoning, she argues, is an active, value-laden practice, evidential reasoning is context dependent and relies on background assumptions, and science is a social inquiry that, at its best, produces social knowledge with contextual objectivity. Her proposal that social interactions are necessary to the management of social values is familiar, drawing as it does on arguments from Mill, Pierce, and Popper, but also novel in providing a more comprehensive theoretical framework than any of her predecessors. Longino takes seriously the social nature of knowledge as it is produced by scientists. She also discusses “what a value-free science might be, why it cannot be, and how we can avoid the paradoxes inherent in more traditional accounts by treating scientific knowledge as social knowledge.”² This project is continued in her 2002 book *The Fate of Knowledge*.

---

¹ Brown.
In what follows I identify the key elements of Longino’s critical contextual empiricism (CCE), including her four CCE norms. I then develop modified versions of these norms in order to make them more comprehensive and more specific. I also clarify Longino’s position on the three predicates of knowledge: content, agency, and warrant. Following this, I reply to a number of objections to CCE raised by Miriam Solomon, Alan Richardson, Alvin Goldman and Philip Kitcher. I aim to offer a description and defence of the key elements of CCE and to provide CCE with greater substantive content and thus enhance its relevance to practical problems. The version of CCE I present, which gives principles of diversity a more central role than the original, will form the basis of the critical arguments in the remainder of the dissertation. The impact of my own interpretations of the theory will be particularly significant in chapters six and seven.

1. **Constitutive and Contextual Values**

It is generally accepted that scientific inquiry is influenced by certain constitutive or epistemic values such as consistency, simplicity and empirical adequacy, which play an important role in guiding scientific inquiry. In addition to these constitutive values, contextual values – individual, social, political and economic values – shape the practices, questions and data, as well as the specific and general assumptions of scientific inquiry. There is extensive empirical evidence of the contextually value-laden nature of science from science and technology studies, sociology of science, and feminist philosophy of science (all drawing on historical and contemporary case studies from science). Given extensive literature documenting the presence of social values in science, contemporary philosophers of science acknowledge that scientific inquiry is not free from contextual values. The philosophical project is to find ways to limit their influence on knowledge.

---

3 Ibid., 4-7. For example, a commitment to making a profit from research is a contextual value that influences the practice of science. Constitutive and contextual values can be positive or negative. So, for instance, if transparency is a positive value in science, secrecy is the corresponding negative value. I discuss positive and negative values further in Chapters Five and Seven when I investigate assumptions about bias and objectivity in medicine.

4 Perhaps the most famous of these are Steven Shapin and Simon Schaffer’s *Leviathan and the Air Pump* on the values in experimentation, Karin Knorr-Cetina’s analysis of decision-making in the laboratory in *The Manufacture of Knowledge*, and Donna Haraway’s investigation into gender bias in primatology in *Primate Visions*. For an overview of many of these arguments, see Brown.
Longino argues that it is not possible to keep contextual values out of science. This is a strong claim to begin with, but she commits herself to something even stronger. She argues, “not only that scientific practices and content on the one hand and social needs and values on the other are in dynamic interaction but that the logical and cognitive structures of scientific inquiry require such interaction.”

In making this argument, Longino goes beyond the more common (and weaker) position that we must tolerate the values that remain in science. Longino makes the stronger claim that values are integral to good science; a science that did manage to succeed at eliminating values would be a greatly impoverished one.

2. **Background Assumptions**

The mechanisms Longino provides for managing values stem from a claim about the underdetermination of theories by data and the mediating role of background assumptions in inquiry. In what follows I will explain these core elements of CCE. Underdetermination is “the in-principle possibility of constructing multiple empirically equivalent, mutually inconsistent theories for any given body of evidence.”

Now an entrenched problem in philosophy of science, underdetermination is primarily an epistemological problem concerning our ability to produce and justify knowledge using only hypotheses and evidence. For any piece of evidence there are multiple competing hypotheses and the choice of hypothesis cannot be fully determined on the basis of the evidence alone. This is because the evidence is consistent with more than one hypothesis, creating a logical gap in scientific reasoning. This gap, argues Longino, is filled by background assumptions.

Longino explains the role of background assumptions in scientific reasoning using an analogy to background conditions for causal interactions. When someone creates enough friction between two dry sticks to produce a spark, we are quick to label the action of rubbing the sticks together the cause of the spark. However, if it were the case that this scenario occurred in an

---

6 Longino, *The Fate of Knowledge*, 127.
7 Other philosophers might call background assumptions ‘principles of inference’, but, according to Longino, this does not capture the contextual (and non-principled) nature of most background assumptions. I will stick with her terminology. Longino, “Essential Tensions – Phase Two,” 260.
8 Philosophers with Humean sympathies may be less inclined to assign causal relationships so quickly. This just brings them to the conclusion of the analogy more efficiently. Humeans may also be inclined to point out that ‘background’ assumptions and ‘foreground’ assumptions are all just part of a complicated set of conditions that
atmosphere radically different from our own – one in which, for instance, there was no oxygen – rubbing the sticks together would not produce a spark. So while we would not normally identify background conditions such as the presence of oxygen as a cause of the spark, it is in fact one of many factors (causes) that make the spark possible. Longino describes background conditions as the “enabling” conditions of causal interaction. Similarly, background assumptions play an enabling role in the formation of a relationship between evidence and hypotheses or theories, even though such assumptions are often neglected in full descriptions of systems.9

Longino argues that background assumptions determine whether a particular state of affairs is considered evidence for a hypothesis. Evidence, for our purposes here, is “an observation, fact, or organized body of information, offered to support or justify inferences or beliefs in the demonstration of some proposition or matter at issue.”10 The same data become evidence in support of different hypotheses because of different background assumptions. In other words, what counts as evidence in a particular domain is, at least in part, determined by context. Given the relational nature of evidence, then, it is not possible to distinguish evidence simpliciter from what is taken to be evidence. The temperature of a patient, for instance, only becomes evidence of the flu in light of background assumptions about the nature of the action of viruses on the human body and physical theories about the action of mercury in glass thermometers. Background assumptions will always play a significant role in the assessment of evidence. We determine what counts as good evidence by means of the assumptions we make.

Once she has established that background assumptions play an ineliminable role in inquiry, concerns about how such background assumptions are justified lead Longino to propose CCE. If background assumptions are to fill the logical gap between hypotheses and evidence, then what justifies the background assumptions? And how do we prevent idiosyncratic social values (such as gender biases) from becoming background assumptions? Since background assumptions cannot be ultimately justified on the basis of further assumptions or on the basis of evidence without regress, and since by their nature they are often invisible to those who employ them, Longino argues for the importance of public criticism of background assumptions. It is people with perspectives and background assumptions that are different from one’s own who are

---


10 Upshur, VanDenKerkhof, Goel, 93.
in a position to critically evaluate one’s assumptions. Traditional accounts of justification, which emphasize particular principles of inference, are still necessary, though insufficient for a full account of knowledge-production. Longino offers a reconceptualization of justification that preserves many of the traditional elements but adds a social dimension. It is no longer possible for a person to produce knowledge in isolation; only when her background assumptions have been subjected to the scrutiny of others can she be said to have knowledge. Without the addition of this social element, underdetermination remains a serious problem. CCE is, in large part, an attempt to solve this problem of underdetermination. More specifically, CCE is an attempt to improve standards of justification in a way that, rather than eliminating values in science, acknowledges them and subjects them to critique.

Longino claims that philosophers have a special role with respect to background assumptions in science. “Part of the work of a critical philosopher of science is to make these assumptions visible and to understand the systematic role they play in scientific inquiry.”

According to this argument, philosophers of science should pay greater attention to the background assumptions in various scientific contexts. Longino is skeptical of the approach to “bad science” in which the critical process begins by identifying suspicious science (research which appears to have been guided by, for instance, racist or sexist values) and ends by showing how such research failed to follow proper scientific standards and guidelines (because ideology interferes with the critical capacity required by scientific research and led the researcher to make unwarranted conclusions). The argument that value-laden science is merely “bad science” runs into difficulty when considering cases in which “good science” – that is, methodologically rigorous research – still manifests value-laden assumptions. I would suggest that many of the most insidious cases of bad science – certainly of bad medical research – appear at first glance to be beyond reproach, as they adhere carefully to currently accepted standards of evidence. The problem lies in the values underlying those standards of evidence.

According to Longino, an approach that claims to have managed values by drawing up and enforcing methodological rules will ultimately allow the deepest influences of social values on science to persist unquestioned. There is significant danger associated with narrow attempts to eradicate the pernicious influences of values in science. This danger arises when the

---

11 The Dualist Staff, “An Interview with Helen Longino.”
12 I discuss this problem in great detail in Chapters Four through Seven of this dissertation.
methodology that is used to distinguish between the good and bad science is itself value-laden, that is, in “those cases where the warrants themselves – that is, the methodological procedures or framing assumptions accepted within a field – are ideologically driven or value laden.” In such cases, shallow investigation into values will only lend the methodologies and background assumptions greater authority. Furthermore, the rules and standards put in place to police values will themselves be value-laden, though their value-ladenness will be unquestioned. Thus, what we need is “a way to block the influence of subjective preference at the level of background assumptions involved in observation and inference, as well as the influence of individual variation in perception at the level of observation.” Longino argues that a social mechanism is needed to perform this function. I will turn now to her proposed strategy for dealing with background assumptions and contextual values in inquiry.

3. CCE Norms

The mechanism Longino proposes for ensuring broad critical debate on background assumptions will be referred to as the CCE norms. According to CCE, scientific communities are objective to the degree that they meet four criteria:

1. *Venues*: there must be recognized avenues for criticism of evidence, of methods, and of assumptions and reasoning.
2. *Uptake*: the community as a whole must be responsive to such criticism.
3. *Public Standards*: there must be shared standards that critics can invoke.
4. *Tempered Equality*: intellectual authority must be shared equally among qualified practitioners.

Collectively, these criteria ensure that the conditions for effective public debate are in place. The point of the first criterion is to ensure public accessibility to the views of others. Recognized avenues for criticism are necessary to ensure that public debate can occur. This criterion is increasingly challenged in the medical context by requirements that researchers enter into confidentiality agreements with private funding agencies. Such confidentiality agreements

---

14 Ibid.
may require researchers to refrain from publishing results without approval or refrain from discussing results with other researchers for extended periods of time. The specialization of medical journals could also be a problem, if it were to mean that critical perspectives were relegated to different journals from those which published original research. Longino stresses the importance of equal space for criticisms in the same journals in which original research is published.

The second criterion ensures that the community not only tolerates, but actively cultivates criticism. If criticisms are leveled, and then ignored, the community fails to be responsive and thus fails the second criterion. Quite often critical papers are written without any indication that the community as a whole has taken note. In my experience, philosophers are particularly good at meeting this criterion. Within the philosophical community, criticism and response is very common, with journals publishing original research side by side with critical evaluations of that research, and conferences often have “author meets critics” panels.17

The third criterion ensures that there are shared public standards to which community members and critics can appeal.18 Shared standards are necessary because they provide a common language or point of agreement from which disagreements and different interpretations can be measured and assessed. Standards are shaped by the cognitive goals and practical aims of a community.

The fourth – and most controversial – criterion ensures the presence of a diverse range of perspectives within a given community. It does so by ensuring “tempered equality” of community participants. This accords equal recognition to perspectives regardless of social or economic power or authority, but does allow for differential recognition of the perspectives of individuals with specialized training and expertise.19

4. Modifications to CCE

While I agree that the four CCE norms do a decent job of protecting the social structures that contribute to effective critical debate – and certainly a better job than the competing social

---

17 This is common at the annual meetings of the Canadian Philosophical Association, for instance.
18 This criterion was originally referred to as shared standards, In The Fate of Knowledge it was renamed public standards. It is my understanding that this is a shift in emphasis rather than in content.
19 Longino, The Fate of Knowledge, 131.
epistemological theories, as discussed in the last chapter – I think it is worth being a bit cautious about what each criterion accomplishes. If we examine the criteria closely, we realize three things. First, the requirement that there are recognized avenues for criticism is not specific enough. The criterion should protect the publicity and transparency of ideas presented within a community. It cannot do this if it only protects venues for criticism without specifying further criteria that those venues must meet. If, for instance, commercial interests were to limit the presentation of research results to the community by restricting publication of the negative results of research, the presence of recognized avenues for criticism will not, alone, bring these hidden results forward. This criterion should protect communities – particularly scientific communities – from secrecy and selective presentation of results. In order to do so it should require that members of the community submit their ideas for consideration in community venues. In addition, it should require that presentation of ideas in these venues occurs under conditions of transparency and full disclosure. I discuss these issues in greater detail in Chapter Seven.

Second, the requirement of shared standards is a good one but needs to highlight the fact that outsiders can engage with a community as long as they share at least one standard with that community. This helps to protect the fourth criterion (cultivation of diverse perspectives, see below) while preserving a sense of unity within given communities. I discuss the details of my views on shared standards later in this chapter.

Finally, the requirement of tempered equality should be replaced with a requirement that communities cultivate diversity. Longino suggests that equality of authority is required to ensure that “every member of the community be regarded as capable of contributing” to any debate.20 It also protects the results of community debate from being secured by “unforced assent”; tempered equality of authority shields inquiry from undue influence from the economically or politically powerful.21 There are a number of problems with this norm. First, there is reason to think that the requirement of shared community standards accomplishes Longino’s goal of ensuring any member of a community has the authority to engage with a debate. Shared standards are the only legitimate basis for inclusion and exclusion of perspectives in a given community, so any further restrictions will be unjustified. Thus the shared standards norm protects the voices of the

20 Ibid., 132.
21 Ibid.
oppressed. Second, the requirement of tempered equality isn’t strong enough. It fails to protect the perspectives of those outside a given community. The norm ensures that, within a given knowledge-productive community, there is tempered equality of authority. But what the norm needs to ensure is that there is sufficient attention to those perspectives outside a given community (provided they meet the shared standards norm already outlined).

Third, the other three norms have specific practical implications which the equality of authority norm, as stated, does not. For example, the importance of recognized avenues for criticism immediately provides community members with a practical task: developing and defending such avenues. While the details of this development are context-specific, they are not difficult to work out. Requirements for peer-reviewed journals or public debates will fall out of this first criterion. The revised first criterion that requires community members to subject their ideas for public discussion, and which requires transparency and disclosure on those ideas, does more to combat political and economic interests than any particular criterion stressing equality of authority. This is because it offers a practical social mechanism for action. But what is thought to be the practical implication of the requirement for equality of authority? It amounts to a demand that community members respect the perspectives offered by others and respond in a collegial manner provided the criticism is advanced by someone who shares some community standards. This requirement merely sums up the other three norms, rather than adding to them. Protection of the community from political and economic interests is not something to be legislated through a special CCE norm. Intellectual independence or integrity is best seen as one of the collective outcomes of the four CCE norms when they are enacted together.

The one element of the fourth criterion worth retaining is its underlying commitment to diverse perspectives. Despite a great deal of discussion about the importance of diverse perspectives throughout Longino’s publications, such perspectives are not adequately protected by the CCE norms. It is diverse perspectives that lead us to objectivity.\textsuperscript{22} We would do better to drop the less practical and somewhat redundant elements of the fourth criterion and focus it on the preservation and cultivation of diverse perspectives. This preserves what matters in Longino’s fourth norm while radicalizing it somewhat. The details of my views on diversity will be discussed in the next section.

\textsuperscript{22} I discuss the specific mechanisms by which diversity contributes to objectivity later in this chapter.
I believe that the modifications I suggest are entirely in line with the goals of CCE. My revised list of CCE norms is thus:

1. *Avenues for Criticism* – there must be recognized avenues for criticism, and these avenues must be *publicly accessible* and require *transparent disclosure* of all relevant information (including competing interests) from those who present their ideas. It must also be a community requirement that all members present their ideas for critical scrutiny if they wish them to be recognized as knowledge.

2. *Responsiveness to Criticism* – the community must be responsive to criticism.

3. *Shared Public Standards* – there must be some shared standards that determine community membership. Outsiders to a particular community are welcome to engage in critical debates as long as they share at least one of the community standards with the target community.

4. *Cultivation of Diverse Perspectives* – communities must cultivate diverse perspectives, that is, the perspectives of those who express strong dissent (explained in the next section).

In what follows I will provide a more detailed account of the third and fourth criteria and the potential tension between the two. Mention of the CCE norms will refer to my modified account from this point forward.

5. **Cultivating Diverse Perspectives**

The cultivation of diverse perspectives is the last of the four revised CCE norms. The preservation of diversity is thought to move communities toward greater objectivity. In what follows I will provide support for the central role accorded to diversity within CCE. Much of the support I provide is not directly from Longino, since she spends little time addressing the specifics of requirements for diversity (her attention is focused on the requirements of equality of authority). Some parts of the views to follow are therefore my own, but I believe they are consistent with her position and helpful for extending CCE to practical problems in the medical context. A clear sense of what diverse perspectives are meant to achieve, and how they do so, is important for the arguments of the sixth chapter.

5.1. **Preliminaries**
Two distinctions will be helpful. First, a demand for greater diversity might refer to the cultural backgrounds of trial participants, the gender or socio-economic backgrounds of researchers, the topics of investigation, the sources of funding, the methodologies, or the theoretical frameworks used in a particular field, to name just a few possibilities. Social epistemologist Miriam Solomon has recently suggested that “diversity is a blunt epistemic instrument,” and that while diversity is good for knowledge, the specific type of diversity does not really matter that much. It is unimportant whether researchers are culturally diverse or a variety of topics are investigated as long as diversity of some sort is playing an epistemological role. I do not want to go as far as this, since I believe this position is indefensible, so let me be a bit clearer on the type of diversity I will defend.

Diverse perspectives increase the likelihood of full critical debate when they highlight dissent. It is this dissent from mainstream or accepted theories that provides the impetus for debate and discussion, which in turn contributes to epistemic strength of theories. I am not interested in weak dissent, which involves mere disagreement with the status quo. Rather, I am interested in those perspectives that give rise to strong dissent; holders of such perspectives pursue, develop, and implement alternative research programs based on alternative methodologies and/or theoretical frameworks. It is strong dissent that I argue is required by CCE.

Second, there are (at least) two different senses of objectivity in common use. On the one hand, objectivity of method has to do with the non-subjective or non-idiosyncratic nature of the criteria used to develop, test, and accept evidence, hypotheses and theories. This sense of objectivity is tied up with concerns about values and bias. On the other hand, scientific data is said to be objective when it is believed to accurately fit with reality. This is referred to as objectivity of data, and is related to issues of scientific realism. In what follows, I show how

---

24 A version of this distinction was proposed by Deborah Tollefsen at the 3rd Annual Episteme Conference (University of Toronto), May 2-3, 2006 in a talk entitled, “Scientific Teamwork: Is there room for dissent?” I have modified it to better fit with the issues raised by this chapter. It is a rough distinction, useful only for highlighting the deep character of dissent discussed in this chapter.
25 Longino, Science as Social Knowledge, 62. Scientific inquiry is often said to be objective in the second sense when it is objective in the first. The relationship between objectivity of method and objectivity of data is more complicated than this, but this general idea is correct. Objectivity of method certainly contributes to objectivity of data, though there may be other requirements unique to establishing objectivity of data depending on one’s ontology.
diverse perspectives contribute to the management of values, or objectivity of method, in science.

5.2. Standard Epistemological Argument for Diversity

It is useful to recall the brief but powerful arguments for diversity offered by John Stuart Mill, which form the core of both Longino’s position and my own. I quote him at length:

First, if any opinion is compelled to silence, that opinion may, for aught we can certainly know, be true. To deny this is to assume our own infallibility.

Secondly, though the silenced opinion be an error, it may, and very commonly does, contain a portion of truth; and since the general or prevailing opinion on any subject is rarely or never the whole truth, it is only by the collision of adverse opinions that the remainder of the truth has any chance of being supplied.

Thirdly, even if the received opinion be not only true, but the whole truth; unless it is suffered to be, and actually is, vigorously and earnestly contested, it will, by most of those who receive it, be held in the manner of a prejudice, with little comprehension or feeling of its rational grounds. And not only this, but, fourthly, the meaning of the doctrine itself will be in danger of being lost, or enfeebled, and deprived of its vital effect on the character and conduct: the dogma becoming a mere formal profession, inefficacious for good, but cumbering the ground, and preventing the growth of any real and heartfelt conviction, from reason or personal experience.  

There are three logical possibilities with respect to the unorthodox or alternative perspectives outlined in any domain of inquiry: such views may be entirely true, they may be partially true, or they may be entirely false. As Mill argued in On Liberty, we cannot deny the first of these possibilities (that unorthodox views may be true) without assuming our infallibility. Since this assumption is not warranted, we must allow that we may be entirely wrong and others may be entirely correct at any given moment. Acknowledgement of our fallibility compels us to cultivate and critically examine all alternative viewpoints on a subject. If, as suggested by the second option, the unorthodox views are partially true – and this seems very likely to be the case much of the time given the history of reform and revolution in science – it is advantageous for us to encourage open discussion of alternative views so that whatever portion of truth they hold may

26 Mill, 64.
be recovered. Finally, it may be the case that unorthodox views are entirely false. It will no doubt be tempting to suppress unorthodox views so as to minimize confusion and to avoid wasting our time and resources. Mill suggests, however, that it is still best to allow such perspectives to be aired and subjected to open debate, since it forces the defenders of the received view to provide careful rational grounds for their position. This increases understanding of the received view and requires that its proponents revisit the motivation, inspiration, reasoning, and experiences underlying their view. It is this reconnection to the passions and reasons behind accepted positions that saves those positions from dogmatism.  

There might be some concern that Mill’s arguments were only intended to apply to freedom of expression in the public realm, that is, to the limits of governmental interference. Science, the objection goes, is a specialized realm with unique characteristics, and arguments about the value of diversity of opinion may not apply in such a realm. In response to this concern, recall that Mill suggests the need for “freedom of opinion and sentiment on all subjects, practical or speculative, scientific, moral or theological.” In addition, Mill writes,

> On every subject on which difference of opinion is possible, the truth depends on a balance to be struck between two conflicting sets of reasons. Even in natural philosophy [science], there is always some other explanation possible of some facts; some geocentric theory instead of heliocentric, some phlogiston instead of oxygen.

Mill’s arguments are not limited to the boundaries of governmental interference; they go beyond the political realm to the epistemological. It is not only morally good or politically advantageous to encourage the development and presentation of diverse perspectives, it also advances our understanding of the world. The central thesis of Mill’s argument is that diversity of opinion is valuable for social, moral and epistemological progress. This position is underwritten by recognition of the inescapable fallibility of human inquiry in all domains, including both politics and science.

### 5.3. CCE Argument for Diversity

---

27 Ibid., 64-5.
28 Ibid.
29 Mill in Lloyd, S404.
In light of the fact that values are pervasive in science, we can extract from CCE a commitment to increasing transparency through disclosure of the assumptions, values and goals of researchers and encouraging critical interaction among scientists not only on the results of research, but on all aspects of the scientific process. CCE attempts to manage values through requirements for transformative critical debate within scientific communities. This focus on the scientific process leads to an account in which objectivity is a characteristic of the practice of a scientific community.

For many, this candid acceptance of values in science (albeit coupled with a management plan) looks like a betrayal of scientific ideals held for hundreds of years. In response, Longino argues that “it is the more frightening prospect of a science continually at the mercy of dominant interests, a science that, under the guise of neutrality, helps create a world to serve those interests,” which should be of greater concern to us.30 Elizabeth Anderson sums this position up eloquently:

Insistence on the value-neutrality of scientists is self-deceptive and unrealistic…Indeed, it is self-defeating: when scientists represent themselves as neutral, this blocks their recognition of the ways their values have shaped their inquiry, and thereby prevents the exposure of these values to critical scrutiny.31

In addition to the standard suggestions for weeding out inappropriate values in research, Longino’s approach highlights the need for attention to diverse perspectives in science. This is the only way we can get at the deepest source of values in science: background assumptions.

Background assumptions shared by all members of a particular community are often invisible to its members. This invisibility can render such assumptions immune from criticism even when criticism is generally encouraged. The presence of diverse perspectives (of the type producing strong dissent) is vital for bringing these assumptions to the attention of members of the community and thus advances the community toward objectivity.

[Background assumptions] do not become visible until individuals who do not share the community’s assumptions can provide alternate explanations of the phenomena without

30 Longino, Science as Social Knowledge, 15.
31 Anderson, “Feminist Epistemology and Philosophy of Science.”
those assumptions...until such alternatives are available, community assumptions are transparent to their adherents.32

Thus, diverse perspectives, which draw the attention of the community to its own assumptions, contribute to the critical evaluation of the assumptions. The views advanced by diverse communities of researchers are an “epistemological resource,” and this resource should not go untapped.33 The presence of diverse perspectives allows researchers to “compare standards and to assess (purported) knowledge produced and accepted in one community in reference to standards proper to another.”34 The only way to achieve this sort of cross-community debate and critical evaluation is through the cultivation of, and open engagement with, communities in which strong dissent is present. This engagement makes possible a thorough critical interrogation of background assumptions and advances inquiry toward greater objectivity. It also encourages critical debate on more superficial assumptions and values such as those present in the results of published research. Longino sums up these arguments on diversity stating:

[T]he greater the number of different points of view included in a given community, the more likely it is that its scientific practice will be objective, that is, that it will result in descriptions and explanations of natural processes that are … less characterized by idiosyncratic subjective preferences of community members than would otherwise be the case.35

Diverse perspectives contribute to the methodological objectivity of scientific inquiry and therefore to knowledge. If we fail to attend to the arguments made by those presenting unorthodox views, “the assumptions shaping our inferences [will be] hidden, [and] we will not see the level at which we might entertain and seriously develop new ideas”36

When accepted knowledge in a community is contested by outsiders, and the community members engage in critical discussion with those outsiders, there is an opportunity to revisit and revise the community’s goals and background assumptions. This process will give rise to more

32 Longino, Science as Social Knowledge, 80.
33 Longino, The Fate of Knowledge, 155.
34 Ibid. “The ideological dimensions of theories of human evolution or of the role of gonadal hormones in behavior, and in general of any theory accepted, are best revealed through comparison with alternatives. Just so, the ideological dimensions of mainstream standards of theory appraisal may be revealed by comparison with an alternative set.” Longino, “Taking Gender Seriously in Philosophy of Science,” 339.
35 Longino, Science as Social Knowledge, 80.
36 Ibid., 225.
objective outcomes, which contributes to the accuracy of scientific research over the long run. “The point in exposing the... biases lying behind certain research theories is not to show that they are false (they might in the end be empirically vindicated), but to make salient the room for alternative programs not based on such biases.”

The process is characterized as one in which knowledge is, “produced collectively through the clashing and meshing of a variety of points of view.” When more time is spent testing, analyzing, and critiquing new hypotheses (rather than extensively developing accepted hypotheses), there is a greater likelihood of accuracy in the end results. Diversity thus increases the objectivity of the methods of scientific research.

6. Shared Public Standards

According to my modified version of CCE, scientific communities produce contextually objective knowledge to the extent that four norms are present: recognized avenues for criticism; responsiveness to criticism; shared public standards; and diversity of perspectives. In what follows, I am particularly interested in the requirement that scientific communities have shared standards, as this requirement may be in tension with the commitment to diversity defended above. In this section, I consider whether a lenient policy on methodological and theoretical diversity leads to a collapse of rigorous standards of evaluation and a disastrous “open-door” policy on membership in scientific communities. In other words, I consider whether it is possible to maintain a commitment to shared scientific standards while upholding a commitment to diversity. One can infer a way of addressing these issues from Longino’s position, though she does not address the issues in much detail. I suggest that as a result of common cognitive goals, scientific communities necessarily share a commitment to empirical adequacy. This requirement protects scientific inquiry from becoming overly inclusive and sets the boundaries of the scientific community as a whole. I then consider whether, once this minimal requirement is met, members of particular scientific communities might still have reason to exclude some evidence, hypotheses, or values from consideration. I develop and specify the conditions that must be met...

37 Anderson, “Feminist Epistemology and Philosophy of Science.”
38 Longino, Science as Social Knowledge, 69.
39 The trade-off is a decrease in fruitfulness as less attention is paid to the careful development of well-established hypotheses and theories. I will discuss this problem further in the final section of the chapter.
40 Ibid., 76-80.
in order for particular outsider communities to be denied a voice in scientific debates, and thus how diversity and rigorous standards in science can be jointly upheld.

6.1. The Tension

The demarcation problem in philosophy of science is concerned with how and where we draw the boundaries of the scientific community and how we determine the principles and standards that will be used to distinguish members from non-members. Various solutions to this problem have been advanced, though demarcation is still a contested topic in the philosophical literature. Demarcating the scientific community from other types of communities is arguably made more difficult when we have an explicit commitment to cultivating and embracing diversity. When diversity – especially that which gives rise to strong dissent – is encouraged and cultivated in the scientific community, it may be more challenging to find common ground upon which to build jointly acceptable community standards. Groups that share few metaphysical or epistemological commitments may be hard-pressed to identify overlap with others, yet we know that common standards are necessary.

Limited resources plague every human endeavor, and the production of knowledge is no exception. Given that time will always be a limited resource, and that material resources (particularly money) are usually in short supply, a lack of standards in science can lead to harm both directly and indirectly. Direct harm comes about when evidence is produced without criteria for judging the value of such evidence, allowing, for example, for the production of fraudulent medical treatments and potentially dangerous ‘cures’. Given the tremendous quantity of pseudo-science produced and the harms that result when the public mistakenly places trust in the results of such pseudo-science, opening community membership to any and all interested parties is deeply problematic. In addition, the epistemic price of doing so is high, as it becomes difficult to arbitrate claims to knowledge. A lack of standards in science also indirectly leads to harm when good scientific research receives fewer of the available resources (as a result of sharing resources with everyone), thus making scientific innovations less likely.

Thus, as Heather Douglas has recently noted, there exists a tension between “shared standards that provide a basis for discussion/agreement and the diversity of participants” in any
given community.\footnote{Douglas, “The Irreducible Complexity of Objectivity,” 464.} Community standards must be loose enough to allow a diversity of perspectives to be openly discussed and debated, and unified (or narrow) enough to permit shared goals and common methods of critique and evaluation. If standards for community membership are set too narrowly this may result in exclusionary practices and a corresponding decrease in the presence of diverse perspectives. As we have just seen, diversity is crucial to the objectivity of scientific inquiry. On the other hand, allowing anyone to be a part of a particular scientific community will mean greater difficulty in finding and agreeing on shared goals and thus greater difficulty in setting and upholding standards – and community standards are necessary. Longino’s contextual empiricism contains one way of settling this tension. In what follows, I outline what I take to be the important features of this solution.

### 6.2. CCE and Community Standards

Three key elements make up critical contextual empiricism. The first (the critical part) stresses the importance of intersubjective debate within and among knowledge-producing communities. This was discussed in the diversity section above. The second (the contextual part) involves recognition of the important role played by background assumptions in knowledge-productive inquiry and the importance of context (particular cognitive goals and practical ends) in shaping how those background assumptions are determined. This was discussed in the background assumptions section above. The third (the empiricist part) involves commitment to a sufficiently modest form of empiricism. While Longino does not make the positivistic claim that all meaningful claims must be empirical, or even the strong empiricist claim that empirical knowledge should be the only basis of knowledge claims, she does commit to a modest position that is at what she calls the “commonsense core” of empiricism, in which “experiential or observational data are the only legitimate bases of theory and hypothesis validation in the empirical sciences.”\footnote{Longino, “Essential Tensions – Phase Two,” 261.} The one condition placed on all communities purporting to be scientific is that they adopt empirical adequacy as a community standard. That is, scientific communities must be responsive to the “truth of the observationally determinable portion of theories or
models." Empirical adequacy requires that there be at least one theory, hypothesis or model under which all the observations can be captured. This is referred to as “saving the phenomena.” There may be several competing theories or models capable of saving the phenomena in particular cases – this is the problem of underdetermination. Nonetheless, the commitment to empirical adequacy is an important common element in any proposed theory or model. One could not, for instance, just selectively throw out data in order to make a certain theory appear better supported. Empirical data, while not the only factor in hypothesis or theory choice, are the “least defeasible bases of hypothesis and theory validation.” Scientists are required to save the phenomena if they are to retain their ability to interact successfully with the empirical world. This does not in itself explain how scientists should evaluate theories or determine community membership, but it does offer a starting point from which to evaluate scientific communities.

Thus, though scientific communities are diverse they are united in their commitment to one shared standard: empirical adequacy. References to “the” scientific community indicate the overlap of all particular scientific communities that have declared allegiance to this standard: “a global scientific community whose members answer to some shared and some not shared standards.” A standard of empirical adequacy follows directly from a cognitive goal of understanding the natural world and a practical end of interacting successfully with that world. The commitment requires that scientists take seriously the empirical results of their interactions with nature. Instances of data-fixing and selective reporting of results run counter to this requirement. Empirical adequacy will be one standard among many accepted within particular scientific communities. But – critically – within any community purporting to be scientific, empirical adequacy must not be overturned or disregarded in favour of other community values. Empirical adequacy must take priority among community values in order for the community to count as scientific. Community members must, at the very least, refrain from fraudulent use of empirical data and have a sincere commitment to discovering new information about the empirical world.

Longino offers some general guidance on dealing with the longstanding problem of demarcation in the sciences. In her evaluation of New Age crystalology and creationism, for instance, she argues that they can be legitimately excluded from mainstream scientific inquiry

---

43 Longino, *The Fate of Knowledge*, 185.  
45 The Dualist Staff, “An Interview with Helen Longino.”
because they fail to meet any of the accepted community standards of scientific communities. Longino declares, “to the extent that [crystalology and creationism] satisfy none of the central standards operative in the scientific communities of their cultures, they indeed qualify as crackpot.” She is referring here to the conspicuous absence of any respect for empirical adequacy within the creationist or crystalologist communities. These pursuits do not count as scientific, in even the broadest sense, because they fail to take seriously the basic scientific commitment to empirical adequacy.

6.3. Local Community Standards and Regulatory Standards

Beyond this shared commitment to empirical adequacy, which results from the cognitive goals of all scientists, there is no static set of community standards common to all communities within the domain of science. Though scientific communities often share some community standards, particular communities are likely to choose a range of different regulatory standards in response to a variety of local factors. Longino writes,

In claiming that public standards are required for a knowledge productive community, I am not claiming that there is a single set of standards that characterize all scientific communities. I’m claiming instead that there is a pool of standards – cognitive, substantive, and practical – that communities draw on in regulating themselves… Different, but overlapping, sets from this pool characterize different communities.

Thus, medical research communities are likely to have community standards that incorporate commitments shared by others within the scientific domain, such as empirical adequacy, as well as standards specific to the cognitive goals and practical ends of medical research, such as mechanism, reductionism, relevance to pressing medical problems, cost-effectiveness and practical value of the treatment for patients. A cognitive community is defined as “any group bound by some set of common goals and shared public standards,” where the standards serve to

---

47 Longino, “Taking Gender Seriously in Philosophy of Science,” 335.
48 “The general family of standards from which those locally adopted might be drawn would include such cognitive virtues as accuracy, coherence and breadth of scope, and such social virtues as fulfilling technical or material needs or facilitating certain kinds of interactions between a society and its material environment or among the society’s members.” Longino, “Subjects, Power and Knowledge,” 112.
regulate the production and evaluation of knowledge. Adherence to community standards is a requirement for community membership and such standards are inferred from the cognitive goals and practical ends of the community. Scientific communities are particularly successful cases of cognitive communities because of their well-defined goals (including a focus on knowledge-production as a primary goal) and rigorous methodologies. Community standards are best understood as principles or values that guide discussion and interaction, rather than as strict or absolute rules. Metaphysical and epistemological commitments form much of the content of community standards.

In addition to community standards, there are also lower-level rules and guidelines governing inquiry that are adopted within particular scientific communities. These tend to have well-defined restrictions and prohibitions and are often created in response to practical concerns such as determining the most appropriate method for answering a certain type of question within a particular domain. The EBM hierarchy of evidence is an example of this type of standard. Such a standard outlines rules and restrictions on methods to be used by the mainstream medical community. A distinction will be useful here: I will refer to the more general, principled standards as community standards, and the more restrictive rule-based standards as regulatory standards. If community standards address the issue of what questions need answering, regulatory standards address how to best answer those questions. This is an important distinction, as those who defend the need for “rigorous standards” in science have a tendency to conflate the two types of standards and thus impose unnecessarily restrictive demands on communities (or on those on the fringes of those communities). Standards of both types are designed, adopted, and revised in response to community goals and in light of challenges encountered when implementing the standards in practice. In this dissertation, I will be interested primarily in those assumptions that shape the content of regulatory standards.

---

49 Longino, The Fate of Knowledge, 145.
50 Most human communities qualify as cognitive communities insofar as they are interested in representing the world accurately in order to achieve their ends.
51 I will describe the hierarchy in Chapter Three and critique it in Chapter Five.
52 This distinction will be helpful in the latter chapters of the dissertation.
53 In the case of both community and regulatory standards, the community dictates what the content should be. The difference is in the content of the standards, not in who proposes the standards, as noted in the definition above. When Longino writes of the need for diverse, open discussion on background assumptions, she is referring to both the global assumptions that shape research within an entire domain and which are the content of community standards and the more specific assumptions that make up regulatory standards.
6.4. Requirements for Inter-community Debate

In order for one community to be considered among those to which another community must pay attention and engage in critical debate, one criterion must be met: the requirement of overlapping standards. Longino argues that some, but not all, standards have to be shared in order for the research produced by one community to claim some right to consideration by those in another. She is speaking here of what I call community standards (since that is her focus throughout her work): “Criticisms and endorsement, as well as the proffering of alternative explanatory models, are made germane to a given community by appeal to some one or more of the standards it recognizes.” Longino suggests that commitment to at least some shared standards is necessary to provide common ground for effective critical debate.

It may be that particular communities are homogenous and have well-defined standards and boundaries, but as long as there are many of these communities, each with its own well-defined standards and boundaries, diversity will still be served. As long as scientific communities are open (and responsive) to critiques from outside their communities (from those who adhere to largely – though not entirely – different standards), they are meeting this requirement.

The goal of critical debate among a variety of communities is not to come to universal consensus, but “to make possible the refinement, correction, rejection, and sharing of models.” To the extent that two communities share some community standards and a common domain, they will be compatible enough to engage in productive dialogue and debate. Longino argues that a community must at least share the same domain as another community, and therefore have some relevance to that community’s concerns, in order to receive attention from that other community. In other words, “alternative theories must be perceived to have some bearing on the concerns of a scientific community in order to have a hearing.” For example, medical researchers would not normally be required to pay attention to research in geology since geologists work in another domain and tend not to share concerns with human health.

Unorthodox views should be attended to when they are critical of the consistency of a mainstream research program, or of the background assumptions that were used to pursue such a

---

54 Longino, “Taking Gender Seriously in Philosophy of Science,” 335.
56 Longino, The Fate of Knowledge, 148.
program. Because various non-mainstream communities share cognitive goals and practical ends, as well as some community standards, with the mainstream, they are in a position to be critical of the assumptions that underlie the regulatory standards of the mainstream. Somewhat ironically, when the members of the mainstream community demand that alternative communities adhere to all mainstream community standards, they are doing damage to their own goals and ends. It is through attentiveness to those who are anomalous by mainstream standards that significant critical discussion about community background assumptions occurs. In these cases, the community can progress toward greater objectivity. I discuss these issues further in Chapter Six.

7. Clarification: Three Senses of Knowledge

“‘Knowledge’ and its cognates are honorific terms, designating success.”\textsuperscript{57} The epistemological literature is replete with variations (and equivocations) on the terms ‘knows’, ‘known’, and ‘knower’. Careful attention to the different ways we understand these terms is vital to a full appreciation of CCE, and will help to avoid common misunderstandings. Longino identifies three independent aspects of knowledge: agency, content and warrant.\textsuperscript{58} I outline each of these aspects in what follows.

7.1. Agency: Who Knows?

Let us begin by considering what it means to say that a particular person knows something. When we are concerned with knowing, we are concerned about the subjects of knowledge and their relationship to some object, or the knower to what is known.\textsuperscript{59} We want to know what it is that makes this person (or these people) know something. We are motivated to ask questions about the relationship between subjects of knowledge and the objects they purport to know because we want to be able to distinguish opinion from knowledge. Most traditional philosophical accounts have been designed to explain this relationship. If you run across an “S knows that P” statement in the epistemological literature (followed by a list of conditions), you are probably looking at a theory that aims to explain who knows in this sense. Traditional

\textsuperscript{57} Ibid., 97.
\textsuperscript{58} I am drawing upon distinctions made in Longino’s 2002 book \textit{The Fate of Knowledge}.
\textsuperscript{59} Ibid., 80.
accounts assume that we are concerned about how *individuals* know. Non-individualists about knowing include holists (who suggest that only groups know), eliminativists (who argue that nobody knows) and socialists (who suggest that interdependent subjects know). As we saw in the last chapter, the most attractive of these positions, assuming we have established an interest in examining the social conditions of knowledge, is the latter, which preserves some of the common intuition that individuals know (which the former two theories cannot) without deferring entirely to individualism. It is thus intuitively appealing as well as likely to be innovative.

7.2. Content: What Do They Know?

When we refer to that which is known – the content or object of knowledge – we draw upon the second aspect of knowledge. We are not referring to a particular person (as in knowing above) or any particular process (as in warrant below). We are attentive to knowledge as a product. Longino suggests that this is often the starting point for philosophical inquiry. Monists believe that there is one correct, consistent and complete account of natural processes. In contrast with this, there are three categories of non-monists: antirealists (or constructivists) who hold that multiple accounts of natural processes can be offered, none of which have a necessary relationship to the real world, eliminativists, who hold that no correct accounts of natural processes are possible (even if there is an independently existing real world), and realists who hold that multiple satisfactory accounts of natural processes can co-exist. The most attractive of these positions is a tempered version of the latter, which allows for the possibility of either pluralism or monism. We do not need to presuppose either monism or pluralism; rather we can leave the question open and see what sorts of empirical results we end up with. According to

---

60 Ibid., 91. These are not exhaustive lists of the possible positions. Any number of nuanced positions are possible those these are the basic divisions. The same goes for the positions identified in the two categories below.

61 That is, philosophers begin by specifying knowledge as a product; driven by a desire to offer a complete account of truth, philosophers begin from the assumption that inquiry aims at truth. Or, if not truth, then some other (single) goal such as utility or agreement. Longino suggests that the disembodied, detached character of knowledge in traditional epistemology is in part a consequence of this starting point. How philosophers characterize knowledge as a product has implications for the characterization of subjects who know, and of justifications (warrants) of knowledge.

62 Ibid., 90.

63 Ibid., 91-2. It is more common for realists to be monists, but a minority of realists are pluralists or, at least, open to the possibility of pluralism.
Longino, we do not need to settle the question of what the content of knowledge will ultimately be in order to get on with the project of epistemology. It is worth noting the monist/pluralist distinction, though, because monism is often assumed and this can lead to restrictions on who knows and how knowledge is justified.

### 7.3. Warrant: How Do They Prove That They Know?

Empirical investigators, in contrast with philosophers, are said to start from an analysis of knowledge-productive processes. This shapes the way they look at the state of having knowledge, as well as what is seen as legitimate content of knowledge. How knowledge is produced matters to whether it is justified or not. “Divination, tea-leaf reading, [and] the dictate of civil or religious authority” are thought to be dubious belief-forming processes and should be excluded from accounts of warrants for knowledge.\(^6^4\) Relativism, whether in epistemology or ethics, is generally thought to be a bad outcome since it licenses arbitrary judgements – often with harmful results. In contrast with relativistic views, there are three main types of non-relativism. The most common is absolutism (sometimes referred to as foundationalism), which upholds context-independent criteria for justification. Eliminativists suggest that justification is either unnecessary or impossible. Finally, contextualists argue that justification relies on the rules and principles shaped by the particular context of inquiry. Debates between foundationalists and contextualists occur over justification, with foundationalists accusing contextualists of collapsing into relativism and contextualists challenging foundationalists to provide grounds for their foundations. CCE is a contextualist theory.

### 7.4. The Traditional Triumvirate

A comprehensive theory of knowledge must address each of these three senses of knowledge. Traditionally there has been a bundle of these accounts of knowledge, which has included: individualism about agency, monism about content, and non-relativism about warrant.\(^6^5\) Longino suggests splitting these commitments apart. Her account adopts non-individualism about agency

---

\(^6^4\) Ibid., 79.

\(^6^5\) Ibid., 89.
(this is the social part of her account), non-monism about content (though she leaves final
determination of this open), and despite diverging on these first two points, she aims to retain a
commitment to non-relativism about warrant. She does this by adopting contextualism rather
than foundationalism. To sum up, Longino’s account of knowledge may be represented as
follows:

Some content A is epistemically acceptable [justified] in community C at time t if A is or is
supported by data d evidence to C at t in light of reasoning and background assumptions
which have survived critical scrutiny from as many perspectives as are available to C at t,
and C is characterized by venues for criticism [and transparency and disclosure in such
venues and requirements that community members use such venues], uptake of criticism,
[shared] public standards, and [a diversity of perspectives].66

7.5. Avoiding Misunderstandings

Alvin Goldman has located Longino’s critical contextual empiricism in the anti-classical
category of social epistemology. This means he does not think Longino’s theory of knowledge
engages with traditional concepts of truth or justification. This classification is mistaken.
Longino explicitly engages with debates over justification (something anti-classical social
epistemologists avoid), and her account is ultimately compatible with traditional epistemology.
Longino describes her work as “not an abandonment of traditional epistemological questions, but
their rearticulation for empirical subjects. This rearticulation does not reject traditional
epistemological concepts.”67 In fact, Longino explicitly rejects a number of competing accounts
of social epistemology precisely because they abandon traditional epistemology.68

In light of the types of knowledge just outlined, we can see why Goldman might have
made this mistake in classifying CCE. It is clear in Longino’s writings that she is not committed
to monism about the content of knowledge: she does not hold that there must be one complete
account of all natural processes. She does not set a unified single goal for inquiry (such as truth).
This rejection of a single aim for inquiry can appear to be an acceptance of relativism (though we
know this is not the case), because the traditional bundles of knowledge have associated non-
monism with relativism. Similarly, Longino’s attention to non-individualism might suggest that
she accepts relativism. Of course, she is a contextualist (non-relativist) about justification, and

66 Ibid., 135.
67 Ibid., 173.
68 Ibid.
actually upholds plausible versions of non-individualism and non-monism, but this point may too often be missed.

Also contributing to this misunderstanding is a distinction Goldman makes in his book *Knowledge in a Social World*. He classifies social epistemologists as either consequentialist or proceduralist. Consequentialists set a single aim or goal for inquiry, ranging from truth (veritism) to agreement (consensus consequentialism) to utility (pragmatism or utility consequentialism). Proceduralists, in keeping with their name, are interested in the epistemic value of processes and practices independent of any particular aim or goal. Goldman claims that while Longino’s account appears to be proceduralist, when examined more carefully it “turns out to have consequentialist contours.” Moreover, Goldman argues that further analysis would uncover an ultimately veritistic rationale for Longino’s social epistemology. Goldman is partly correct and partly mistaken in this characterization. Longino’s position is compatible with truth as one of the goals of inquiry. The mistake here is again in assuming that Longino is tied to a monist position about the content of inquiry. In fact, as we’ve seen, Longino has explicitly built an openness to pluralism into her account. She is not interested in settling the debate over the goals of all inquiry, and wants to leave room for pluralism about these goals. So Longino’s proceduralism is ‘infected’ by attention to goals only insofar as she leaves open the possibility for a variety of goals in the context of inquiry. Her account may appear to be somewhat veritistic in character because truth is at least one of the possible goals of inquiry. But because the goals of inquiry are not specified by Longino’s account, arising instead out of the particular needs of different knowledge-producing communities, her account takes as its starting point procedures of justification (warrant) rather than the (single) aim of inquiry. To be fair to Goldman, insofar as particular communities adopt goals (such as truth), Longino’s account can appear to be consequentialist in nature. But this is a product of the way people choose to develop knowledge-productive communities in the real world rather than a feature of the account.

### 8. Objections and Replies

Longino’s work has been well received by many philosophers of science, social epistemologists and feminists. It has also been subject to a number of critiques. I will turn now to some of the

---

69 Goldman, *Knowledge in a Social World*, 78.
more common concerns with CCE. I supply my own responses to these critiques, and in doing so I draw upon my version of CCE as described above, though I believe my responses to be in line with arguments that Longino would make.

8.1. Failure to Engage with the Messy “Real” World of Science

In a review of Longino’s (2002) book *The Fate of Knowledge*, Miriam Solomon and Alan Richardson decry Longino’s inattention to “the messy details of the real world in which science gets done,” suggesting that her normative account “may be neither possible nor necessary for scientific progress.”70 This overarching criticism has a few distinct elements, some of which gain more traction than others. I will move from the more straightforward to the more difficult objections.

8.1.1. Intuition

First, Solomon and Richardson agree that Longino is correct in recognizing the need for a normative account of scientific inquiry, but are concerned that she seems to be developing her account solely on the basis of intuition – which has a “poor track record in philosophy of science”.71 They suggest that Longino’s claims to be generally naturalistic are misleading since she has no significant commitment to incorporating insights from scientific practice into her theory.

In reply to this concern, it is first worth noting that Longino clearly states that her account is only naturalistic insofar as it incorporates a descriptive understanding of human subjects and their contexts and goals at the outset. It is not naturalistic in any other, more robust sense. In fact, she states that “filling in the details between the empirical conditions of knowledge production and the ideal conditions of success is the job of philosophical theory.”72 It is true, however, that Longino relies on intuition, reasoning and explication of concepts in order to develop her normative account. But it is not clear how this is a robust criticism of Longino. It amounts to a

71 Ibid., 213.
72 Longino, *The Fate of Knowledge*, 10. For further discussion on Longino’s position on naturalism, see Chapter One of this dissertation.
claim that Longino’s account is not naturalistic and therefore it is bad. But this is just a
restatement of the naturalist’s position along with a critical evaluation of anyone who is not a
substantive naturalist. Longino is in good company with most other social epistemologists in her
philosophical methods. Her account would seem to stand or fall based on whether there are good
reasons to be concerned about the limitations of a thoroughly naturalistic approach to
epistemology. These limitations are highly debated. Moreover, it is not as if philosophers
propose an intuitive idea and then dogmatically stick with it at all costs. The epistemological
theories advanced by philosophers like Longino are open to general discussion and critique in the
analytic epistemological community. Traditionally, this sort of epistemological debate has been
both rigorous and critical: the community is hardly one to adopt a view without question. So
Longino’s account of justification – her suggestion that underdetermination in the sciences leads
to pervasive values in inquiry and her proposal of four norms to guide inquiry – is available for
general scrutiny. The criticism that philosophers rest on intuitions is misleading in light of the
extremely rigorous community evaluation that occurs in epistemology. By Longino’s own
standards, the epistemological community is knowledge-productive. That it is not scientific
should not come as a surprise, since philosophy is not a science.

8.1.2. Empirical Evidence

Second, in a related concern, Solomon and Richardson point out that Longino does not offer an
empirical evaluation of her approach. Longino fails to evaluate whether communities that adopt
the CCE norms produce better quality results than other communities. Longino could, for
instance, perform empirical tests comparing communities with the CCE norms to communities
with different norms in order to see whether there is a difference. Solomon and Richardson
believe this sort of evidence is needed. This would require the development of a way to measure
the extent to which the CCE norms are being followed, and then a clear sense of what outcome
the communities are aiming for. Neither of these will be a straightforward task. There is no
reason to think that Longino would be opposed to this sort of evaluation, though it may be very
difficult to get around the circularity involved in setting the primary outcome of investigation. As
Solomon and Richardson admit, Longino has indicated a willingness to do this sort of evaluation
(in primatology). The fact that she has not yet done it is not an indication of resistance to the idea.

**8.1.3. History of Science**

Third, Solomon and Richardson argue that Longino’s account fails because it cannot be applied to cases in the history of science. For instance, in Longino’s book she argues that the persecution of witches in the early modern period, though characterized by certain rules and standards, failed to be subjected to a high level of critical scrutiny. The practices also failed to retain a commitment to accurately describing the empirical world. As a result, the persecutors did not have knowledge – and certainly did not have scientific knowledge. Drawing on work by Lorraine Daston, Solomon and Richardson argue that Longino’s account of what was going on in these cases is too simplistic. They claim that historical evidence indicates that witch ‘theorists’ were actively involved in critical debates with one another (thus fulfilling the requirement for critical debate) and members of the scientific community and were committed to a certain version of accuracy (though not quite the one we usually recognize). Thus, they did satisfy the CCE norms and, arguably, even upheld the commitment to empirical adequacy. Accordingly, they should be considered knowledge-productive communities, and possibly even scientific communities.

Longino is interested in avoiding this outcome. In order to do so she is forced to tighten her four norms – for instance requiring a more robust social criticism. She must claim that communities of witch theorists were insufficiently open to criticism because they ignored concerns raised by interested parties, including most obviously the women being persecuted. But if she does this it appears as though “nearly all early modern science” will fail to meet the new higher standards. Consequently it appears as though persecutors of witches and early modern scientists will stand or fall together as legitimate scientists on her account.

Solomon and Richardson also suggest that the difficulties encountered with historical cases extend to more contemporary cases. They offer an example of one case – rejection of Darwinism in the early years – where scientists failed to meet the criterion of tempered equality

---

73 Daston, “Marvelous facts and miraculous events in early modern Europe.”
(or, more precisely, respect for diverse perspectives given shared standards). They did so by failing to recognizing the work of less-known scientists (such as Darwin at the time).

In reply to these concerns, Longino has two options. First, it is worth examining how responsive communities are required to be in these circumstances. The Darwin example is helpful here. Longino is not blind to the large body of science studies literature that indicates the importance of a certain degree of conservatism in scientific practice. There is no reason to think that “reasonable responsiveness” to challenges requires that scientists immediately give up their theories when confronted with alternatives. What is more often the case is that some elements of a theory are problematic while others are not and the choice between theories is far from straightforward. In these most common situations, responsiveness means being aware of criticisms and the weaknesses in one’s account and responding to criticisms leveled by others. It means not turning a blind eye. Recall that Longino is open to pluralism and is a firm advocate of diversity of theories and methodologies in science and so is not trying to suggest that responsiveness be interpreted so radically that a slight breeze knocks over any theory.

Second, Longino can bite the bullet and admit that much of early modern science was not as scientific or as productive of knowledge as we would like to think. This reply would be supported by claims she makes about the need for us to “do better in the twenty-first century than our forebears did in the eighteenth.” Though she is referring to philosophers (not scientists) when she says this, I think the sentiment suggests a willingness to admit the significant failings of science – whether today or in the past – and a desire to hold knowledge up as an honorific term that is not ascribed lightly. In fact, we do think it was problematic that Darwin was ignored in favour of more socially acceptable scientists in his day in spite of the high quality of his work. The importance of respecting diverse perspectives is demonstrated, not undermined, by this example. Longino is offering a normative account of ideal science. Of course it will be possible to find instances of scientific inquiry that fail her standards. This does not immediately signify the need to jettison her account. In fact, if all the account did was explain exactly why we are right about everything already it would not be much of a normative account.

8.2. Failure to Capture What is Distinctive About Scientific Evidence

Longino, “Reply to Philip Kitcher,” 577.
Alvin Goldman is concerned about a lack of specificity in Longino’s account. He writes,

"Doesn’t more have to be said about the types of evidence distinctive to science (experimental evidence, presumably) and how, specifically, the evidence is deployed (methods of inference)? These dimensions are not adequately captured by the abstract and otherwise unconstrained requirement of interactive criticism."\(^{76}\)

Longino’s account, while more detailed than the general accounts offered by predecessors such as Pierce, is still quite general. It offers four norms and not too much more. Is this enough?

Goldman is correct in pointing out that more needs to be said on this point, though perhaps unfair in suggesting that this will be as straightforward as he implies and in charging Longino, or other advocates of CCE, with the task. Given Longino’s commitment to local, contextual knowledge in specific communities, she cannot give much more detail on these points without specifying the context further. But what Goldman seems to want are general scientific methods and accounts of evidence. If Goldman would accept local, contextual responses to his questions, then it is not clear why Longino is the one who should be providing such responses. According to her account it is the members of communities who set the standards and rules and defend those standards against others. As a result, if you want to say more about methods of inference and standards of evidence you need to engage in discussions with those community members. This is an interesting project, but one for empirical (not exclusively philosophical) investigation.

**8.3. Improper Use of Liberal Democratic Principles**

Philip Kitcher has argued that Longino’s interest in socializing science is laudable but that she injects the social element at the wrong point in inquiry.\(^{77}\) Kitcher famously argues for the democratization of science policy. He argues that the public should play a role in setting the research agenda in science. He also argues for the importance of social discussion and debate on the results of research. So the social element plays a role before and after research, so to speak. Longino argues that we need to socialize what is in between these two stages of inquiry – the method itself. This is the key point of contention between their accounts, which otherwise share

---

\(^{76}\) Goldman, “Knowledge and Social Norms,” 2149.

\(^{77}\) Kitcher, “The Third Way.”
a commitment (or, at least, an openness) to pluralism, an emphasis on the context-dependence of science, a recognition that social interests need to be addressed in science, and a sort of liberal democratic spirit. Longino playfully refers to Kitcher’s position as “socialism lite” because, like Goldman, Kitcher maintains a very weak account of the social and because social interactions play only a contingent role in knowledge-production (rather than a necessary role as is the case in Longino’s account).

Kitcher argues that Longino sees the need for a more robust role for the social because she confuses two problems: the Millian Problem and the Interest Problem. The Millian Problem arises when the exclusion of certain groups of people from scientific discussions leads to a decrease in number of available hypotheses in science. Greater diversity in the scientific community has drawn attention to the Millian Problem and some possible solutions. For example, female primatologists revolutionized the field by proposing new hypotheses about the interactions of primates that relied less on assumptions about the importance of the dominant male. To use a medical example, the inclusion of alternative medical researchers in discussions within the mainstream medical research community may increase innovation and new hypotheses about contested illness such as Chronic Fatigue Syndrome (about which very little is known).

The Interest Problem arises when the interests of certain subgroups of the human population are reflected in the type of research that gets done (the questions asked) and the sorts of hypotheses accepted. For example, if the majority of scientists are middle-class Caucasian men (as was historically the case), medical research is likely to reflect their health concerns and focus on conditions that differentially affect that demographic, such as heart disease, high cholesterol, and impotence. To use a recent example, the 90/10 problem in medical research (90% of the resources are directed to diseases that affect 10% of the population) is a result of policies that allow researchers to pursue lucrative projects over pressing ones. According to Kitcher, democratization can help solve these problems, but in different ways. The solution to the Millian Problem is to insist on diverse communities as well as responsiveness to criticism from those outside one’s particular community. The solution to the Interest Problem is a democratization of the research agenda (of the sort proposed by Kitcher). Longino clearly identifies and addresses the Millian problem (in fact, much more thoroughly than Kitcher does). The concern seems to be that in focusing on the Millian problem she fails to address the Interest
Problem. But Longino does address the Interest Problem, if in a slightly different manner than Kitcher. Longino argues that the presence of idiosyncratic interests can only be managed if the assumptions underlying research are subjected to wide critical evaluation by diverse community members and outsiders. Given the overlapping knowledge-productive communities of which most people are a part, public criticism of research agendas will be possible (for instance, in the form of critiques of certain assumptions that profit matters more than health). To be fair, Kitcher goes much further in developing the specifics of such a solution, but the general compatibility of Kitcher’s and Longino’s democratic tendencies would seem to suggest the possibility for a reconciliation on the Interest Problem. Longino would not be doing her account any disservice by agreeing to a role for the public in influencing research programs. The focus of her account is elsewhere – in the careful development of solutions to the Millian Problem. Finally, as she points out, even if we are successful at democratizing the research agenda (that is, ensuring that we pursue questions of significance), there will still be a need for democratization of scientific practice in order to ensure that we most accurately answer those questions. Longino’s epistemological project is distinct from Kitcher’s political project.

8.4. Dogmatism

Do communities exhibiting the qualities required by CCE produce knowledge? Goldman presents a test case for Longino’s account. He describes a religious community in which members form beliefs on the basis of the dictates of a sacred text. Because different interpretations of the text are possible, critical discussions and debates over the correct interpretation of the text are common and even encouraged. Community members adjust their interpretations in light of insights by others in accordance with publicly available standards. According to Goldman, this community appears to adhere to the four (original) CCE norms, but we would not want to classify this community as a knowledge-producing community.

Similar cases are outlined in Longino’s own work, and she responds to them by pointing out that that communities have to be open not only to the criticisms of community members but to the criticisms of members of other communities who share at least some of the same public standards. Insofar as religious communities are closed to the criticisms of other communities – that is, insofar as they are dogmatic in their commitment to certain texts even when confronted
with good evidence in favour of other texts, or against all texts – they fail the requirement of
responsiveness. Further, on the basis of my discussion on diversity, such communities would
have to cultivate strong dissent. But it is unlikely that they do so. This leads to Goldman’s core
concern: the same sort of dogmatism seen in religious communities is present and vital in
science. He characterizes the commitment to scientific methods as dogmatic because
“researchers aren’t invited to challenge those [scientific] methods when they submit their
research papers.” Further, even if scientific methods are challenged in some venue “the
challenge is assessed by appeal to logic and mathematics. And aren’t those standards dogmatic in
some sense?” His point is that Longino’s requirement that communities remain open to
criticism of their methods (and, I would add, cultivate diverse perspectives) is too strong and that
no knowledge-producing community will be adequate to the task.

In the medical context, physicians do criticize research methods – that is what critical
essays and opinion sections of medical journals are for and why there are usually journal sections
specifically designed for these sorts of debates. The commitment that members of particular
scientific communities have to methods and methodologies is not dogmatic, and insofar as there
is a tendency toward dogmatism because researchers do not tend to challenge the methods used
in their community, this is a significant and serious failing. Goldman is suggesting that a certain
amount of non-vicious dogmatism is necessary to scientific inquiry. But what would scientists
have to fear if they were to be open to debates over the details of their chosen methods?
Underlying Goldman’s critique is a belief that if you are not dogmatic in defending your
standards they will collapse. But if you have a good defence of your standards there is no reason
to think they will not hold up to critical scrutiny, and be all the stronger for surviving such
attacks. Standards are dogmatic when people fail to have reasons for them, and fail to give
reasons to others who inquire about them. Scientific standards are not, or at least should not be,
dogmatic in this sense. A commitment is not dogmatic just because it is not repudiated in favour
of another. It is dogmatic when it is upheld even though there are no reasons in its favour, or at
least none significantly better than those offered for alternatives.

9. Summary

78 Goldman, “Knowledge and Social Norms,” 2148.
79 Ibid.
In this chapter, I outlined the key elements of CCE and a novel interpretation of the four CCE norms. I argued for the importance of diverse perspectives to ensure the ongoing critical evaluation of background assumptions. I then considered whether there is a tension between methodological and theoretical diversity and rigorous scientific community standards and argued that a balance can be struck between these competing values. This balance depends on shared, and indefeasible, commitment to a standard of empirical adequacy within any community purporting to be scientific. This one shared criterion provides the basis for collaborative critical interactions between communities. Finally, I defended CCE against a variety of criticisms and reiterated the importance of critical evaluation of background assumptions in the pursuit of knowledge. In the next few chapters I will be drawing on these lessons as I describe and critically evaluate the standards of evidence proposed by evidence-based medicine.
Chapter Three
Evidence-based Medicine

“Dad always thought laughter was the best medicine, which I guess is why several of us died of tuberculosis.”

Surely, if medicine is to be based on something, evidence is a good choice. It is certainly better than “vehemence-based medicine” or “eminence-based medicine”, as some of the more sarcastic defenders of evidence-based medicine (EBM) point out. The ubiquitous EBM movement, now adopted in health care contexts around the world, relies heavily on this initial impression. EBM has gained at least some of its popularity from the intuitively obvious nature of its name as well as its apparently innocent and widely accepted goals. It is the motivations and assumptions behind the movement that I will be describing in this chapter. There are now many different interpretations of EBM available. Of these, I will be focusing on EBM as it has been advanced by the members of the EBM Working Group at McMaster University, since the historical record indicates that their pronouncements are at the core of the movement. I begin with a brief historical overview in order to contextualize the EBM movement. I then describe the unique features of EBM. Finally I outline the substantive background assumptions underlying EBM. The background assumptions identified in the final section of this chapter will be systematically critiqued in chapters four and five.

1. History and Development of Evidence-based Medicine

1.1. Rationalism and Empiricism

---

1 Handy, 21.
2 Isaacs and Fitzgerald, 1618.
3 Note that these terms are not used in the standard philosophical senses. Rationalists in medicine, for instance, do not only reason from first principles. They are actually quite empirically-oriented as they emphasize the importance of empirical investigation into basic mechanisms of disease. The designation ‘rationalist’ was likely picked to highlight the role of reason in this approach. Empiricists in medicine are thought to be interested in ‘whether something works’ regardless of causes or mechanisms. Again, the use of the terminology does not correspond to classic philosophical accounts of empiricism. In fact, this whole discussion might be more helpfully cast as a debate between empiricist approaches to medicine at different levels. While empiricism (in the philosophical sense)
Modern medicine has inherited two competing approaches to the care of patients. One of these, traced back to Hippocrates in ancient Greece, is loosely referred to as the rationalist approach. Rationalism in ancient Greek medicine emphasized the importance of uncovering the mechanisms of disease. Medical doctors could, on the basis of their understanding of physiology, anatomy and other basic sciences, identify problems and reason through to the effects of various treatments on patients. This rationalist approach was advanced by the Hippocratic physicians for many centuries and posited single causes as the source of illness and disease. This meant a diagnosis of imbalance in the four humours (blood, phlegm, yellow bile and black bile) and treatment designed to restore the body’s ideal balance.4 The later empiricists, in contrast, developed an approach to medical practice that eschewed theoretical reasoning in favour of observations of patients. They developed their practices through careful observation of cases and cumulative expertise on a number of cases. The primary interest of empiricists was in choosing the best treatment for a condition, rather than understanding the “first causes” of disease. Preferred empiricist methods of treatment were pharmacology and surgery.5

For the vast majority of the twentieth century, the ascendancy of modern medicine in North America (and concurrent devaluation of otherwise popular systems of alternative medicine) coincided with the rise of rationalism in medicine. The highly influential Flexner Report in the United States in 1910 argued for an increased focus on the basic sciences as a vital part of medical education. The tremendous influence of the report in the West led to additional attention to physiology, anatomy, pathology and microbiology in medical schools. In effect, the report challenged physicians to understand the basic mechanisms of disease. As a result of the changes in medical schools over the twentieth century, by about 1980 the following description of affairs was offered, “the orthodoxy of modern medicine is rationalist; a large majority of physicians within academic medical centers and in practice are subspecialists who are experts in a particular set of diseases and focus on particular organ systems or diseases.”6 Speaking of the historical dominance of laboratory science in hospitals, Kerr White has described the dominant

---

4 Newton, "Rationalism and Empiricism in Modern Medicine.”
5 Ibid. This version of historical events is simplified in order to draw out the rationalist/empiricist distinction.
6 Ibid., 304.
mentality colourfully as follows, “[D]iseases have single causes, and they are mostly ‘bugs’. What I call the Big Bug Hunt was under way: we have got to look for bugs everywhere. Meanwhile, the hygienists who looked at the environment, including the social scene, poverty, economic conditions, and occupational hazards, were cast aside by the biomedical establishment.” Also cast aside were epidemiologists and medical researchers conducting trials on groups of patients. The return to an empiricist approach to medicine, however, came about relatively quickly and with great force. As Warren Newton puts it, “for all its rhetoric of novelty, Evidence Based Medicine represents a counter-revolution of traditional empiricism, draped in modern clothes of statistics and multi-variate analysis.”

1.2. Clinical Epidemiology

From approximately the late 1960s to the 1990s, (with significant growth in the late 1980s) physicians in Canada, the United States and Europe began to shift their attention to a new approach to medicine that was based on the increasingly popular methods of clinical epidemiology. The new movement was thought to remedy a perceived over-reliance on laboratory (basic science) research as a resource for medical decision-making. Epidemiological methods were traditionally within the purview of public health but work by a number of scholars, notably Alvan Feinstein at Yale University, sought to adapt these methods for clinical practice – hence, clinical epidemiology. Feinstein’s book Clinical Judgment (1967) provided much of the groundwork for the movement by outlining the need for greater systematization, classification and consistency in medical diagnosis and treatment. The first comprehensive textbook of the new field, Clinical Epidemiology: The Essentials, emerged in 1982. Finally, Feinstein’s Clinical Epidemiology, in 1985, had a significant impact on many practicing physicians who were attracted by his desire to define, classify and quantify the various elements of clinical practice. In effect, Feinstein called for a taxonomy of medical practice. Many of Feinstein’s original suggestions were ignored, but the commitment to using data obtained from populations

---

7 White, quoted in Daly, 42.
8 Newton, “Rationalism and Empiricism in Modern Medicine,” 314.
9 Ibid., 302. See for example: Sackett, Clinical Epidemiology and Feinstein, Clinical Epidemiology.
10 Feinstein, Clinical Judgment.
11 Fletcher, Fletcher, and Wagner.
12 Feinstein, Clinical Epidemiology: The Architecture of Clinical Research.
of patients to guide decisions at the bedside remains the core of clinical epidemiology today. Clinical epidemiology was touted as a new “basic science of clinical medicine,” designed to replace the Flexnerian reliance on the laboratory sciences.

Clinical epidemiologists upheld a commitment to studying populations and using the knowledge gained to guide decisions at the level of individual patients. This was driven by a frustration with the uncertainties of clinical decision-making, dissatisfaction with traditional approaches to medical care and a desire to separate out good from bad medical practices. The first department of Clinical Epidemiology and Biostatistics was established at McMaster University in Canada, under the leadership of David Sackett. It was the members of this department, including, among others, Brian Haynes and Gordon Guyatt, who formed the Evidence-based Medicine Working Group.

1.3. Motivations

The EBM movement arose for countless reasons, each of which is grounded in the social, historical, economic and political contexts of modern health care. I will mention three contributing factors here. The first variable was the incredible growth in laboratory research in the biomedical sciences over the twentieth century. From biochemistry to genetics, laboratory research has expanded significantly and this has spurred research in a variety of health-related fields. These advances, when coupled with vigorous efforts on the part of medical professionals to become more scientific (from the Flexner Report through to pharmacogenetics), led to increased attention to the biomedical sciences. Second, the growth of clinical research, which was aided by developments in statistics in the early part of the century, led to the creation of thousands of new medical journals. Physicians who were trained in basic sciences appeared ill-equipped and often, as a result, ill-motivated to stay on top of the massive quantity of research.

13 Contributions by Danish scholar Henrik Wulff are also worth noting. Wulff’s book *Rational Diagnosis and Treatment*, first published in 1973, echoed many of the sentiments raised by Feinstein, and he later developed a philosophical analysis of clinical medicine in *Philosophy of Medicine: An Introduction*. Particularly in the latter book, Wulff offers some useful insights into the motivations, and also the shortcomings of approaches to clinical decision-making (such as clinical epidemiology) that neglect the humanistic elements of health care. These sentiments are echoed by Jeanne Daly in her book.

14 Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 
(of highly variant quality) published every day. They lacked the statistical knowledge and the critical attitude necessary to tackle the evaluation of such research. The advent of the technological age only exacerbated this problem, as it allowed greater and more efficient access to these journals, through a variety of databases (MEDLINE, for instance).

Third, a number of surveys were produced, indicating that patients with similar symptoms were receiving different treatments depending on the particular physician they visited. This was true even in the case of illnesses where fairly conclusive evidence was present to indicate a particular treatment choice. This “troubling” lack of consistency among physicians reflected two important facts. Physicians were continuing to provide treatments (for example anti-arrhythmic agents such as flecainide and encainide, penicillin for the flu, or shaving and enemas for childbirth) long after they had been proven harmful or unnecessary, and physicians were failing to prescribe treatments even when such treatments had been subjected to considerable testing and were widely regarded as the best available treatment for a particular condition. The potential for significant harm to patients in these circumstances is clear. EBM was developed as a response to these and other concerns and was a natural outgrowth of the popular clinical epidemiology movement that had been occurring in the West since the 1960s.

2. EBM Version One

Evidence-based medicine is the application of the techniques of clinical epidemiology to the problems encountered at the bedside. In the early 1990s, The Journal of the American Medical Association (JAMA) published a series of articles in which the members of the Evidence-based Medicine Working Group introduced and carefully outlined a new approach to medical teaching and practice. In characteristically grandiose language, the group pronounced:

15 “The difficulties that clinicians face in keeping abreast of all the medical advances reported in primary journals are obvious from a comparison of the time required for reading (for general medicine, enough to examine 19 articles per day, 365 days per year with the time available (well under an hour a week by British medical consultants, even on self reports.” Sackett et al. “Evidence Based Medicine: What it is and What it Isn’t,” 71.
16 Tanenbaum, “‘Medical effectiveness’ in Canadian and U.S. health policy: the comparative politics of inferential ambiguity,” 518.
18 Daly.
A new paradigm for medical practice is emerging. Evidence-based medicine de-emphasizes intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision-making and stresses the examination of evidence from clinical research. EBM requires new skills of the physician, including efficient literature searching and the application of the formal rules of evidence.\(^\text{20}\)

The “former paradigm” was based on the following assumptions about medical decision-making:
1) Unsystematic clinical experience is a good basis for medical decisions, 2) Reasoning from pathophysiologic principles is a good basis for medical decisions, 3) Physicians possess the skills to evaluate the results of clinical research trials, and 4) Expertise and clinical experience are a good source of knowledge on which to base guidelines for clinical practice.\(^\text{21}\) Unsurprisingly, the “new paradigm” disputes these assumptions, one by one. The stated assumptions of the new paradigm are:

1) Experience and intuition are necessary but not sufficient for medical decision-making.
2) Reasoning from pathophysiologic principles is necessary but not sufficient for medical decision-making.
3) Systematic, reproducible and unbiased observations (such as those found in clinical trials) are a good ‘base’ for medical decision-making.
4) Rules of evidence are necessary in order to critically assess the results of clinical trials.\(^\text{22}\)

These assumptions amount to claims that clinical research should be the basis of medical decisions and that new rules, principles, and skills are needed to support this approach.

Recognizing that evidence in medicine comes in many different forms, EBM proponents introduced various hierarchies of evidence (notably those for diagnosis, prognosis, and treatment) designed to help physicians determine the relative value of evidence produced by different research methods. The hierarchy for medical treatments was the centerpiece of the new movement. This hierarchy of evidence was designed to aid physicians in their day-to-day medical decision-making. The primary role of the physician, as proclaimed by initial EBM

\(^{20}\) Ibid., 2420. “Intuition” (in this context) refers to a subtle decision-making method that draws upon years of professional experience and is generally inaccessible to new practitioners. It is often context- and even patient-specific.

\(^{21}\) Ibid., 2421.

\(^{22}\) Ibid.
enthusiasts, was to keep on top of, and critically appraise, current research on a given disease or medical issue, and then to apply the results of this critical investigation to individual patients.

3. The Hierarchy of Evidence

The difficulty of achieving consensus among medical experts led to the creation of the first hierarchy of evidence. David Sackett originally proposed the idea of ranking evidence on a scale as an objective method for resolving disputes among physicians at consensus conferences. Consensus conferences had a tendency to stall once various experts had presented the evidence for their preferred position. What Sackett did was aid the members in developing a method for comparatively assessing evidence. Case reports were graded lower on the ranking, while randomized controlled trials held the top spot. Later developments added meta-analyses of RCTs above single RCTs and the details of the middle-levels were worked out in greater detail.

The evidence hierarchy was designed to reflect the methodological strength of scientific studies. It is assumed that better evidence on this scale is less likely to be infected by bias, more likely to correctly attribute causal powers to a particular treatment and more generalizable. The Oxford Centre for Evidence-based Medicine offers the most well-established version of the hierarchy. Because this version was produced by members of the original EBM Working Group including David Sackett, Sharon Straus and Brian Haynes, I will use it as the focus of my analysis. The hierarchy for therapy, prevention, aetiology and harm ranks research methods as follows:

- Systematic review of RCTs
- Individual RCT
- All or none
- Systematic review of cohort studies

23 The *Users’ Guide to the Medical Literature* (also edited by members of the EBM Working Group) has offered the following version of the hierarchy: N-of-1 randomized controlled trial; Systematic reviews of randomized trials; Single randomized trials; Systematic review of observational studies; Single observational study; Physiologic studies; Unsystematic clinical observations. This hierarchy has the n-of-1 trial at the top, unlike most other hierarchies in the medical literature. This is because of efforts by editor Gordon Guyatt to draw attention to the n-of-1 design. Guyatt and Rennie, *Users’ Guide to the Medical Literature*, 7. It is otherwise very similar to the Oxford version used here.

24 There are multiple competing hierarchies on offer these days. Ross Upshur has collected seven versions of the hierarchy for therapy so far. Though there are some differences, and these differences do matter in practice, most hierarchies share the same general structure and I intend my analysis to apply to any version though I focus specifically on the most well-known hierarchy.
Variations of the hierarchy offered by other agencies tend to follow this same basic pattern. That pattern is:

**Experimental Trials**
**Observational Studies**
Anecdotal Evidence, Bench Research, Physiologic Principles

Systematic reviews are preferred over single trials (of whatever design) and experimental methods are preferred over observational methods. While the EBM movement is ostensibly committed to a broad definition of evidence as “any empirical observation about the apparent relation between events,” the hierarchy implicitly offers a more limited account. In spite of some gestures indicating the importance of an inclusive account of evidence, EBM proponents go on to restrict the scope of good evidence through the hierarchy: “The hierarchy implies a clear course of action for physicians addressing patient problems: they should look for the highest available evidence from the hierarchy.”

According to EBM, a physician searching for a treatment for a particular patient should formulate a question, do a literature search for relevant studies, and consult the evidence hierarchy in order to determine the quality and strength of the results obtained. Once a full critical appraisal has been performed, he or she can then weigh the particularities of the

---

25 Oxford Centre for Evidence-based Medicine. I have simplified the descriptions of each of the levels for clarity. The rare “all or none” trial occurs when all patients died before the treatment became available but with the treatment some now survive, or when some patients once died but now none die with the treatment. In terms of other popular hierarchies, the U.S. Preventive Services Task Force offers this version: “Level I: Evidence obtained from at least one properly designed randomized controlled trial; Level II-1: Evidence obtained from well-designed controlled trials without randomization; Level II-2: Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group; Level II-3: Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled trials might also be regarded as this type of evidence; Level III: Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.” U.S. Preventive Services Task Force. This hierarchy was adapted from one produced by the Canadian Task Force on the Periodic Health Examination and is frequently used in practice.
27 Ibid., 8.
individual case against the determined objective strength of the evidence. The evidence hierarchy was created in large part to assist physicians in weeding out biased research from high quality research. EBM advocates do not make any claims about the complete or perfect objectivity of any particular types of studies. They would be the first to say that they have simply evaluated the relative rigour of different types of medical evidence, and that some evidence fares better in this evaluation. The final result of this evaluation is that “the randomised trial, and especially the systematic review of several randomised trials, is so much more likely to inform us and so much less likely to mislead us.”

EBM is thought to promise certainty of decisions, transparency of reasoning and a “scientific” approach to medicine.

4. Epidemiological Terminology

It will be helpful to have a basic understanding of the most common research methods before continuing any further. In what follows I briefly outline the most common experimental and observational designs. I also introduce some of the general epidemiological terminology that will come up in the next few chapters.

4.1. Experimental Designs

So-called “experimental” trials are contrasted with observational studies. In experimental trials, investigators actively divide the participants into groups. The action of intervening serves to distinguish experimental trials from observational studies. There are two common types of experimental trials: RCTs and n-of-1 trials. The RCT is a trial design whereby subjects, who are similar at the outset, are assigned by random allocation to intervention and control groups. It is otherwise identical to prospective cohort studies in its causal inferential structure (cohort studies are discussed below). The focus in the EBM literature, and in this dissertation, is on phase-three RCTs, because it is at the phase-three stage that treatments – usually drugs – are assessed at the suggested dosage in a large number of people suffering from the condition or disease. In other

---

29 Though widely used, I do not think this distinction holds up under scrutiny. Why, after all, is random allocation of participants into groups the only type of intervention that counts as a true active intervention? The sorts of interventions present in, for instance, case studies do not seem to matter here for some reason.
words, these are the last trials before a drug receives approval for general marketing and use.\textsuperscript{30} The n-of-1 trial was developed by Gordon Guyatt and colleagues in the late 1980s in response to concerns that the traditional RCT fell short of providing guidance for particular individual patients.\textsuperscript{31} It was first used to investigate treatments for a patient suffering from chronic asthma.\textsuperscript{32} The n-of-1 trial is, essentially, a trial of therapy in an individual patient. The n-of-1 uses the standard RCT method but uses a sample size of one. It is a single-patient trial, in which the patient begins with one treatment and then switches to a control (treatment or placebo) after a pre-determined amount of time.\textsuperscript{33} This can go on for as long as desired with further treatments, or back and forth between one treatment and control. The idea is that the patient acts as his or her own control. Just as in standard RCTs, subjects and physicians can be blinded to the order of the treatments (blinding is explained below).

\subsection*{4.2. Observational Designs}

Observational studies are thought to be less trustworthy than experimental trials. In observational studies, participants are either already grouped (as in diseased groups within case-control studies) or choose their own grouping. They are not actively divided into groups by investigators (though they may be recruited into the study by investigators, and thus be selected from the general population). It was observational studies that established the efficacy of, for instance, insulin in patients with diabetic ketoacidosis and hip replacement for patients with debilitating hip osteoarthritis.\textsuperscript{34} The most common types of observational studies are cohort studies and case-control studies.

Cohort studies can be either prospective or retrospective. Prospective cohort studies follow exposed and non-exposed groups over time (with respect to some suspected causal factor), and compare disease rates during the time period. Retrospective cohort studies do the

\begin{itemize}
  \item Phase one trials are performed on healthy human subjects and are often used to establish safety and dosage levels; phase two trials are performed on small groups of subjects who suffer from the disease or condition and investigate efficacy, phase three trials are larger versions of the phase two trials, often on thousands and even tens of thousands of patients. Phase four trials, which are not carried out as often as they should be, track the effects of treatment in the general population once the treatment is widely available. Phase four research is vital to establishing the generalizability of trial results and to examining long-term effects since phase-three trials are often of short duration.
  \item Guyatt et al., “Determining Optimal Therapy: Randomized Trials in Individual Patients.”
  \item Daly, 87.
  \item RCTs with “cross-over” design may also switch patients between treatments.
  \item Guyatt and Rennie, \textit{Users’ Guide to the Medical Literature}, 7.
\end{itemize}
same, but make use of historical records rather than following people forward from the present time. Retrospective cohort studies differ from case-control studies in that they begin by grouping the historical records of subjects according to suspected health factors, rather than beginning with patient who have particular diseases. A cohort study on smoking would follow smokers and non-smokers and compare, for instance, rates of lung cancer in the two groups.

Case-control studies start with a group of subjects who are suffering from a disease (“cases”). The investigators then retrospectively compare the exposure (to some suspected causal factor) of these diseased subjects with the exposure of people who did not contract the disease (“controls”). For instance, a case-control study might investigate whether subjects with lung cancer were exposed to more or less cigarette smoke than subjects in a similar demographic (age, gender, etc.) who did not develop lung cancer. Challenges arise in choosing the appropriate control group, since many different control groups are possible depending on your hypothesis. Finally, case studies are presentations of interesting or unusual individual cases and case series are collections of case studies.

4.3. General Terminology

Efficacy is a measure of the effect of an intervention under ideal conditions. RCTs are relatively good at establishing efficacy of certain treatments. Efficacy is also often referred to as the internal validity, or integrity, of a trial. Effectiveness is a measure of the effect of an intervention under usual conditions. Intention to treat analysis, where subjects are analyzed according to the group they were assigned, regardless of whether they completed the trial, aids in establishing effectiveness. Pragmatic trials (discussed in Chapter Six) can also improve the ability of an RCT to establish effectiveness. Effectiveness is also often referred to as the external validity, or generalizability, of a trial.

Another common term encountered in the epidemiological literature is blinding. In a single-blind trial, the subjects do not know to which group (treatment or control) they belong. In a double-blind trial both researchers (physicians) and subjects are unaware of which subjects are in each of the groups. In a triple-blind trial, the statistician interpreting the results is also blinded to the groupings. There are interventions to which experimenters cannot be blinded (such as
surgical interventions). There are also interventions to which subjects cannot be effectively blinded (such as lifestyle interventions). Effective blinding requires diligence.

Meta-analyses combine the results of various single trials that address related hypotheses. They have the advantage of bringing together data from a number of different primary sources. The most common meta-analyses are those of RCTs. The sample size – the number of subjects studied in a particular trial – is usually much larger in meta-analyses because the data from individual trials under consideration are pooled together.

Finally, a note about the distinction between method and methodology as the terms appear in this thesis. The tools of scientific investigation are its methods. These are the means by which scientific inquiry is carried out. Broadly speaking, a method is the procedure used to attain an end. More precisely, in science a method is “a special form of procedure or characteristic set of procedures employed (more or less systematically) in an intellectual discipline or field of study as a mode of investigation and inquiry, or of teaching and exposition.” 35 The metaphysical, epistemological and practical principles and procedures that determine how such tools are applied are properly referred to as its methodology. Methodology is a branch of epistemology which investigates the methods used in a particular field. In contemporary usage, it involves “the study of the direction and implications of empirical research, or of the suitability of the techniques employed in it; (more generally) a method or body of methods used in a particular field of study or activity.” 36 A methodology “describes and analyzes methods for reaching a specific goal.” 37 There is a tendency to use methodology to mean method (as in “the methodology of this study was a RCT”). This is an improper use of the terminology and it glosses over an important distinction between theory and practice. My concern in this thesis is methodological. In order to accurately describe the methodology of medical research, it will be necessary to look at the particular methods used by researchers and the ways in which they are differentially valued. This allows for the sort of systematic evaluation of medical methods that is necessary to this investigation.

5. Assessment of Evidence in Three Stages

EBM recommends the assessment of clinical research evidence in three stages. I will outline these stages in what follows.

5.1. Validity

According to proponents, the hierarchy of evidence provides a ranking of trial “validity”.[38] How, then, is validity assessed by physicians? The EBM handbook suggests that physicians ask the following sequential questions in order to determine validity:

1) Was the assignment of patients to treatment randomized?
2) Was the randomization concealed?
3) Were the groups similar at the start of the trial?
4) Was follow-up of patients sufficiently long and complete?
5) Were all patients analyzed in the groups to which they were randomized?

Some finer points:
6) Were patients, clinicians and study personnel kept blind to the treatment?
7) Were groups treated equally, apart from the experimental therapy?[39]

Note that assessment of validity begins with three questions related to randomization and that question five assumes randomization has taken place. Presumably question five is worded the way it is because physicians following the rules have jettisoned all non-randomized trials by that point. This emphasis on randomization will be discussed in Chapter Five.

5.2. Importance

Once trial validity has been assessed, the “importance” of the trial is determined. At this stage, physicians ask two questions:

1) What is the magnitude of the treatment effect?
2) How precise is the estimate of the treatment effect?[40]

---

[38] This account of validity bears no relation to standard accounts of deductive validity in logic.
[39] Straus et al., 117.
[40] Ibid., 123.
Note the narrow scope of importance here. There are no questions about clinical significance, but rather assumptions about the magnitude and precision – or narrow confidence interval – achieved by a study. I will discuss these sorts of statistical assumptions in Chapter Five.

5.3. Relevance

Finally, physicians determine whether the evidence is applicable to their patient. They are advised to ask:

1) Is our patient so different from those in the study that its results cannot apply?
2) Is the treatment feasible in our setting?
3) What are our patient’s potential benefits and harms from the therapy?
4) What are our patient’s values and expectations for both the outcome we are trying to prevent and the treatment we are offering?\(^{41}\)

Note the optimism in the first question. Physicians are not asking whether their patients are similar to those enrolled in the trial they are reading, but rather whether they are so different that the results are inapplicable. The authors go on to stress that significant differences usually will not be found and that physicians can generalize on the basis of RCTs without too much difficulty. I will discuss the generalizability of clinical research in Chapter Four. Physicians who follow this extended set of questions carefully when assessing medical evidence are supposed to be able to make an informed recommendation about the appropriate treatment for a particular patient.

6. The Reaction

Reactions to the introduction of EBM were swift. While some of the initial responses were cautiously optimistic, there was also a strong critical reaction from many physicians. Editorials and articles began to appear in medical journals voicing concerns that the art of clinical proficiency and conscientious individual judgement were being lost under the weight of meta-

\(^{41}\) Ibid., 132.
analyses of RCTs. Charges of “cookbook medicine” and concern about the diminished focus on expertise, clinical experience and judgement provoked a response from the members of the EBM Working Group.

7. EBM Version Two

In 1996, several of the authors of the original declaration of EBM published an article entitled, “Evidence-based Medicine: what it is and what it isn’t,” designed to clarify (and tone-down) some of the more contentious claims from the initial formulation. In this oft-quoted article, Sackett and colleagues offer a more thoughtful and carefully worded description of their approach to medicine:

Evidence-based Medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.

This latest definition reflects an attempt on the part of EBM advocates to honour the art as well as the science of good medical practice. The new, more integrative EBM explicitly recognizes the importance of clinical expertise, judgement, intuition and patient values:

External clinical evidence can inform, but can never replace, individual clinical expertise, and it is this expertise that decides whether the external evidence applies to the individual patient at all and, if so, how it should be integrated into a clinical decision.

The change in tone here is palpable. Instead of calling for a revolution, EBM proponents are looking to reconcile the “former” and “new” paradigms of medicine. Or, at least, that is how this appears on the surface.

Response to the latest version of EBM has, once again, been mixed. Some physicians find the idea of basing medical practice on the best available evidence to be beyond reproach -

---


even “obvious”. Others direct their attention to the precise meaning and interpretation of the terms conscientious and judicious in the EBM definition. Still others continue to write in to medical journals with concerns about the shortcomings of the new approach, including: the persistent “grey zones” of clinical practice, the primacy accorded to RCTs in the evidence hierarchy, the absence of discussion on the role of values in medical decision-making, and the tendency to downplay the individual and complex nature of the patient-physician interaction. Defenders of EBM maintain that these criticisms are “misguided”, “misunderstandings”, or even “clearly invalid”. They maintain that there is no reason for physicians to fear the newer, more inclusive version of EBM. Uptake of EBM by physicians has been slow, despite widespread attempts to encourage the use of EBM guidelines.

8. The Background Assumptions of EBM

EBM is a movement premised on a particular account of good evidence. As was discussed in the last chapter, data becomes evidence only in light of background assumptions. When we speak of evidence, we have either some explicit or implicit idea of the framework within which some particular data becomes evidence: “Evidence itself does not recommend its own interpretation.” All evidence is really evidence that meets context-specific requirements: “Nothing can count as evidence for anything except relative to a host of auxiliary assumptions; and the strength with which a body of evidence supports a hypothesis can never be higher than the credibility of these auxiliaries.” How is it, then, that EBM is able to specify simply “good evidence” through the evidence hierarchy? This is because the background assumptions have been filled in. Good evidence in the hierarchy is evidence that meets assumptions outlined by early EBM advocates.

Standards of evidence are not self-evident. Evidential standards are created by identifying common characteristics of good evidence identified across disciplines (it often, for instance,

---

46 Zarkovich and Upshur, “The Virtues of Evidence.”
47 Naylor, “Grey Zones of Clinical Practice: some limits to Evidence-based Medicine.”
50 Laupacis, “The Future of Evidence-based Medicine,” 7A.
51 Longino, Science and Social Knowledge, 215.
52 Upshur, “Seven Characteristics of Medical Evidence,” 94.
increases the likelihood that a given hypothesis is true), but such standards are also designed with specific attention to the contextual features in which the evidence is produced and utilized. In medicine, the EBM standards of evidence are an attempt to impose a particular view of good evidence, and therefore of appropriate treatment, on the practice of medicine. The designers of EBM made a number of assumptions. First, they made assumptions about the nature and goal of medicine as a science. Second, they made assumptions about clinical decision-making. Third, they made assumptions about the need for a hierarchy of evidence. I will identify and describe these assumptions below. Critical discussion of each of these assumptions will follow in the next two chapters.

8.1. Medicine Should Be More Scientific

One of the members of the EBM working group, Gordon Guyatt, originally proposed the name “scientific medicine” to describe the new movement. Because it seemed to imply that physicians were otherwise unscientific it was heavily critiqued by his colleagues and was later jettisoned in favour of evidence-based medicine. The desire of early EBM enthusiasts was to create a more scientific approach to medicine. The EBM Working Group writes, “systematic attempts to record observations in a reproducible and unbiased fashion markedly increase the confidence one can have…in the efficacy of treatment”. While clinical experience and instinct are still valuable, “one must be cautious” with them because they can be “misleading.” The assumption is that physicians should be more confident in their clinical decisions if they stem from systematic scientific evidence rather than from unsystematic clinical experience. This first assumption ties medicine closely to science. Good evidence is characterized as unbiased, systematic and quantitative. It isolates clear causal relationships and is generalizable across populations in a law-like manner. Personal experience and intuitions fail to meet these standards. The value placed on these qualities reflects the authors’ belief that medicine would be improved by becoming more scientific. This is supplemented later in the paper with the final assumption that

---

54 There is extensive debate within epistemology and philosophy of science on the necessary and sufficient conditions – if in fact there are any – for evidence generally. Even the claim that evidence increases the probability that a particular hypothesis is true has been contested. See Achinstein for a full elaboration of this debate.
55 Presumably similar concerns have arisen with the new name!
56 Evidence-based Medicine Working Group, 2421. Emphasis added.
57 Ibid.
“physicians whose practice is based on an understanding of the underlying evidence will provide superior patient care.”

EBM advocates also assume that physicians’ confidence levels (which are meant to be a reflection of better patient care) should be tied to the degree to which evidence is properly scientific.

### 8.2. Clinical Research Evidence Should Form the Basis of Medical Decision-making

EBM proponents assume that clinical research evidence should form the basis of medical decision-making. This is in contrast with decision-making based on authority or pathophysiology.

#### 8.2.1. Not Authority

According to EBM advocates, the way of the past with respect to medical decision-making involved an appeal to authority and wisdom (in the form of accumulated years of clinical experience). The way of the future (EBM) “puts a much lower value on authority”. They say, “the underlying belief is that physicians can gain the skills to make independent assessments of evidence and thus evaluate the credibility of opinions being offered by experts.” In order to use EBM in clinical practice, physicians have to have a good understanding of the strengths and weakness of various clinical research methods (i.e. an awareness of the evidence hierarchy) and the ability to formulate good questions, search the literature, critically appraise the quality and relevance of the information, and convey the findings to the patient. This is essentially a solitary process, with no need for input from authoritative experts or guidelines.

#### 8.2.2. Not Pathophysiology

Knowledge about basic mechanisms of disease is insufficient as a foundation for clinical decision-making. Pathophysiologic reasoning (tracing possible causal mechanisms of disease

---

58 Ibid.
59 Ibid.
60 Ibid.
61 Ibid.
based on accepted scientific theories) is contrasted with outcomes research (experimental and observational clinical research), with the former more often leading to “incorrect” and “inaccurate” clinical predictions. The assumption is that outcomes research is less likely to lead to error in clinical decision-making. The focus of physicians is no longer whether we have a plausible biochemical or physical explanation for a particular treatment, but whether there are studies showing that patients recover on a basis of the treatment. How something works has taken a backseat to whether something works. To use the language proposed at the outset of this chapter, this reflects a shift away from rationalistic approaches to more empirical approaches in medicine.

8.3. A Hierarchy of Evidence is Necessary

“Understanding certain rules of evidence is necessary to correctly interpret literature on causation, prognosis, diagnostic tests, and treatment strategy.” The third assumption is that physicians need rules in order to correctly evaluate current medical evidence. Rules of evidence, according to this assumption, should be designed by experts (epidemiologists and statisticians, for instance) and then followed and applied by physicians. The practice of ranking different types of evidence (as in the evidence hierarchy) is one that is thought to follow from scientific standards. Note that this is not just the claim that it is helpful to be able to distinguish, for instance, good from bad RCTs or better from worse cohort studies. This is an assumption about the necessity of ranking these methods against one another so that a review of the literature will produce one, hopefully decisive, answer.

9. Summary

In this chapter I outlined the history of EBM from its origins in clinical epidemiology to the shift to an ostensibly more integrative version in recent years. I outlined the key features of the account, as proposed by members of the EBM working group, and I identified the three main background assumptions of the movement. I will be looking in greater detail at the justifications

---

62 Ibid.
63 Ibid.
offered for these assumptions in the next two chapters, since “even if scientists are not concerned in their daily practice with directly justifying the assumptions that structure their reasoning, philosophers concerned with the nature of scientific knowledge must be.” The historical record of standards of evidence in medicine indicates that there has been a shift to empirical approaches in medicine in the last half century. Given this shift to empiricism within medicine, it is particularly appropriate that critical contextual empiricism is the framework I am using to understand and critique current standards of medical evidence.

---

64 Longino, The Fate of Knowledge, 127.
Chapter Four

Critiquing the Background Assumptions of Evidence-based Medicine: Part I

Values filter into medical research at two recognized levels: the social level and the individual level. In the first case, for instance, funding agencies may influence the type of research that gets done, who does it, and whether it is submitted for publication; in the second, researchers may deliberately doctor trial methodologies, selectively present results, and allow their personal biases to influence the outcome or presentation of evidence.¹ In addition to the familiar ways in which values influence medical science, critical contextual empiricism (CCE) draws our attention to the values that form the guiding background assumptions of inquiry. According to CCE, these assumptions are the most pervasive and yet least recognized sources of contextual values in science.² In what follows I will critically evaluate two of the background assumptions of EBM identified in the previous chapter: the assumption that medicine should be more scientific and the assumption that medical decision-making should be based on clinical research evidence.³ With respect to the first assumption, I argue that: the characterization of science underlying the drive to make medical research more scientific is inadequate; scientific approaches to medical practice have not yet been shown to provide better patient outcomes; and medical practice is committed to principles of shared decision-making and individualized care which are in tension with the demands of science as characterized. With respect to the second assumption, I argue that while clinical research evidence was meant to replace authority in medical decision-making it has become a new form of authority, and that pathophysiologic reasoning, and for that matter, a number of other factors, are as “basic” to clinical decision-making as research evidence.

1. Should Medicine Be More Scientific?

¹ Lemmens.
² Longino, Science as Social Knowledge, 60.
³ I will be evaluating the third assumption – that a hierarchy of evidence is necessary – in the next chapter.
Proponents of EBM aim to make medicine more scientific. There are two distinct steps to this process. First, they have to ensure that medical research is sufficiently scientific. Second, they have to tie medical practice to medical research. If medical research fails to be scientific enough, there is not much hope for medical practice. If medical practice is too artistic and humanistic, and thus disconnected from research, medicine as a whole will remain unscientific.\(^4\)

### 1.1. Medical Research

#### 1.1.1. The Received View of Science

In order to ensure that medical research is sufficiently scientific, the EBM working group had to come up with an account of “good science” and reorganize medical research to fit that account. The prevalent view reflected in popular culture is that medicine should be a science like any other. This is the received view on the relationship between medicine and science. That is, medicine is situated alongside physics, biology and chemistry in the lexicon of science. Advocates of this position hold that there are laws, theories, hypotheses, and experiments in medicine in much the same way as in other sciences, and that medicine is really a collection of specialized human sciences including for instance: molecular genetics, biochemistry, physiology, anatomy, pathology, clinical epidemiology and so on. According to the received view, theories are “formalized axiom systems that (with the aid of boundary conditions and initial conditions) deductively organize specific subject areas in science, exhibiting empirical laws and individual occurrences as logical consequences”.\(^5\) On this account, science is characterized as: empirical, systematic, objective, quantitative, and able to isolate simple causal relationships and produce universal (or at least probabilistic) laws or generalizations.

#### 1.1.2. EBM and Science

\(^4\) Whether or not medicine is a science, it will still be appropriate to evaluate it according to CCE. Despite the fact that Longino tends to write about science, the scope of her project is not limited to the critical evaluation of scientific communities. Rather, she intends CCE to be useful in understanding and critically evaluating knowledge-producing communities generally. Longino writes, “The argument of this book is directed explicitly at scientific knowledge. There are, however, reasons to see it as encompassing knowledge in general, or at least empirical knowledge in general.” Longino, *The Fate of Knowledge*, 2.

\(^5\) Schaffner, 67. Schaffner is describing, not endorsing, this view.
I argue that advocates of the EBM movement are included among the proponents of the received view. Certain stereotypically scientific characteristics derived from science textbooks and popular culture, and drawing on the received view outlined above, were assumed without critical debate and then imposed on medical research. The assumption underlying EBM is that medicine needs to become scientific with the characteristics just listed. Many of these assumptions found their way into the evidence hierarchy, as we will see in the next chapter. Unfortunately, the received view of science is highly problematic according to philosophers of science. I turn to these concerns now.

There is a significant body of literature demonstrating that the received view of science is untenable when applied in certain domains. Biology and medicine, for instance, are not easily made to produce universal laws and results via deductive logic, or isolate simple causal relationships. This position, advanced in various forms by J.J. Smart, Philip Kitcher and Kenneth Schaffner, suggests that there are important differences between the biomedical sciences and the hard sciences of physics and chemistry. In his book *Discovery and Explanation in Biology and Medicine*, Schaffner identifies ways in which theories and laws in biology and medicine are not the same as theories and laws in physics. He refers to theories in the biomedical sciences as “theories of the middle range” because they fall somewhere between the universal mechanisms of biochemistry and the universal generalizations of neo-Darwinian evolution. Schaffner argues that features of biological and human systems determine this difference. Biology and medicine are characterized as complex, “vigorously interlevel”, “less linear”, “more reticulate,” not productive of universal claims in the traditional sense, unable to “fully specify and control for all relevant initial conditions,” and are said to draw upon an “overlapping series of interlevel temporal models” rather than laws. Theories in these areas consist of a cluster of models and their consequences. Schaffner is concerned primarily with bench science in his descriptions, but the view applies equally well to epidemiological research.

This characterization of biology and medicine is argued to be more accurate, more flexible and more fertile than the received view. Theories of the mid-range are still capable of all the important epistemological features of physical and chemical theories, in that,

---

6 Ibid., 64.
7 Ibid., 65.
8 Ibid., 87-89, 98, 125.
9 Ibid., 95.
[They] are testable and have excess empirical content, they organize knowledge in inductive and sometimes even deductive ways, they change and are accepted and replaced in accordance with rational canons, they have a generality that applies outside their original domains (though this generality is frequently partial and complex), and they are applicable for prediction and control in crucially important areas such as … health care delivery.  

Rather than concluding, on the basis of differences between, for instance, biology and physics, that biology fails to be scientific, Schaffner and others argue that we need a broader account of science. According to advocates of this position, medicine is a science, just a different and unusual kind of science. Medical researchers adhere to critical scientific principles such as prediction, empirical experimentation and generalization and are thus scientific enough to merit the association.

I propose an even stronger critique of the received view of science than is offered by Schaffner. Unfortunately, because Schaffner assumes that physics is the model science and is concerned only with characterizing and assigning legitimacy to the *differences* found in biology and medicine, his account fails to consider more revolutionary possibilities. A better characterization is to suggest that none of the sciences can be taken as paradigmatic in the way physics is usually taken to be, and to recognize that each of the different sciences employs different tools in order to gain knowledge about the empirical world. We do not have real or core sciences and other sciences; we have a range of sciences. And even in traditionally paradigmatic sciences, we cannot apply simplistic ideas about theories and laws. This argument draws support from arguments in the philosophy of physics literature in recent years which indicate that it is not appropriate to talk about universal laws and pure deductive inference in any realm of science – even physics.  

This revelation exposes the misunderstanding underlying the attempt to find universal medical laws. Moreover, it suggests the need to pay attention to contextual differences in various knowledge-productive realms rather than to find ways of factoring those differences out. The medical domain is dominated by questions of specific individual effects. Attention to these questions and the goals of medicine is critical to the success of any knowledge-productive activity in that domain. We have much to lose if this point is forgotten:

---

10 Ibid., 120.
Characterizing as ‘legitimate’ biomedical theories only those theories that fit the analysis more appropriate for physics and chemistry would impoverish existing biomedical science and would methodologically misdirect ongoing research.\textsuperscript{12}

Medicine is a science unlike physics, but is no less a science as a result.

The claim that EBM is more scientific than other approaches to decision-making in medicine gets tossed around a lot in the literature without much careful attention to the scope and content of science.\textsuperscript{13} Medical research should be thought of as scientific because members of the medical research community share a commitment to empirical adequacy, and to certain principles of empirical evaluation. However medicine should not be assessed according to the similarities and differences between it and physics or chemistry. The proponents of EBM have attempted to make medicine more scientific, and in doing so they have borrowed inappropriately from caricatures of the basic sciences. They have attempted to impose demands for law-like generalizations, quantified results, unbiased methods, and single-level causal phenomena in a context where, as Schaffner demonstrates, these are not always possible or appropriate. As I noted above, Schaffner has argued extensively, and persuasively, for the legitimacy of different characterizations of biology and medicine as sciences. Biological and medical theories, for instance, are conceptualized in relation to the specific and unique ways in which they discover, test, explain and justify facts about the empirical world. Attempts to impose inappropriate characterizations of science on medicine obscure these important differences. So while the goal of making medical research more scientific is admirable, the particular account of science imposed on medicine was, in the case of EBM, inappropriate and reflects an outdated understanding of science. This does not preclude the development of a more accurate account of science, in line with my more radical reading of Schaffner’s conclusions, and a strengthening of the relationship between medical research and this more accurate version of science. Further investigation into this possibility is warranted, though beyond the scope of this project. I discuss some of the specific implications of adopting stereotypically scientific characteristics in Chapter Five when I interrogate the assumptions underlying the evidence hierarchy.

\textsuperscript{12} Schaffner, 126.
\textsuperscript{13} Daly, 1.
1.2. Medical Practice

I turn now to the second step: connecting medical research and medical practice. Sciences are all partly knowledge-seeking endeavors and partly applied sciences. It is too easy to say that, for instance, domains like medicine are applied sciences and thus separate from the pure sciences in their attachment to the production of useful technologies. All scientific inquiry is pervasively influenced by social values, interests, expectations and goals. We cannot separate pure inquiry into the empirical world from the applications that drive inquiry in all fields. Recognition of the partly knowledge-productive and partly technologically-productive nature of inquiry today allows us to properly assign responsibility to researchers and all interested stakeholders and community members, including the public. Though no domain of science is immune from practical expectations, some particular domains might be further removed from such expectations than others. Thus while all science contains knowledge-productive and technologically-productive tendencies, some are weighted more heavily to the production of technological interventions.

Medical practice is probably closer to the applied end of the spectrum than medical research. Medical practice is problem-driven and usually aims not for general knowledge but for specific, individually appropriate applications of knowledge to particular patients. Medical practice does not, primarily, aim to expand existing knowledge, but to apply it in order to solve real-world problems, most notably the relief of diseases in individuals.\(^1\) This is not unlike the relationship between theoretical physics and engineering, or theories in ecology and ecosystem management. While theoretical frameworks are developed and hypotheses generated and tested within the more-theoretical domain (for instance physics or ecology), the results of these investigations are put into practice in the more-practical domain (for instance engineering or ecosystem management). The relationship is complex, however, in that the needs and goals of the practical realm inform the research projects of the theoretical realm. Further, engineers and ecosystem managers know better than to directly apply the results of theoretical investigations in

---

\(^1\) The situation is undoubtedly more complicated than this. Physicians receive new information from every patient (thus building up their knowledge of the differences in presentation of various diseases and increasing their ‘clinical expertise’), and thus medical practice itself is partly knowledge-productive. This is captured in my description of medical practice as closer to the technological/applied end of the spectrum (because this still contains a knowledge-productive element). Recall that while this means medicine is partly scientific, I am using my revised-Schaffnerian account of science here (not the received view).
the practical domain. A great deal of careful re-contextualization (for instance, adding real forces such as friction to equations in engineering) is necessary in order to bridge theory and practice.

The desire to make medicine more scientific seems to require a closer relationship between medical practice and medical research than was traditionally the case. This is the argument made by EBM proponents. In order to tighten the connections between research and practice, EBM proposes a series of clinical practice rules (outlined in Chapter Three) that advise physicians to consult the research literature whenever possible in the care of individual patients. This is, in a sense, designed to re-contextualize the theoretical results in a similar manner to the re-introduction of friction to equations in engineering. In the earliest formulations of EBM, physicians were to apply the results of research in practice, and though later versions talked about integration of research with other factors, the desire to connect research and practice persists. In what follows I will take a closer look at some of the reasons why the particular drive to bind medical practice to medical research is problematic.

1.2.1. Patient Outcomes

Even if tying medical practice to medical research (where medical research is sufficiently scientific) were to make medical practice more scientific by association, it is not at all clear that this would ensure better patient outcomes in medical practice. Empirical evidence of this connection is lacking, and it is not at all clear why we would assume that tying medical practice to medical research would be good for patients. This is commonly raised as an ironic problem for EBM: there isn’t any empirical evidence to indicate that EBM is a more successful approach to individual patient care. EBM advocates rely on common sense and expert assessments to suggest that they are on the right track.

EBM advocates are currently in the process of collecting evidence on whether EBM leads to better patient outcomes. So far, evidence indicating the effectiveness of EBM is limited to the results of observational research. This is clearly a matter of some consternation for the advocates of EBM. In the most recent edition of the EBM handbook, Straus and colleagues lament the challenges of producing evidence for evidence-based medicine. They point out that in order to evaluate EBM they need to define a precise intervention, an appropriate dose, and a clear outcome. Unfortunately, these details are hard to work out. The intervention is “difficult to
define”, it is “unclear what the appropriate ‘dose’, ‘formulation’, ‘frequency’, or ‘route’ should be”, and it is difficult to determine which outcomes to measure since EBM produces a “wide range of outcomes”, some of which are “harder to confirm” than others.\textsuperscript{15} In addition, the authors make a point of noting that “learners have different learning needs and styles, and these differences must be reflected in the educational experiences provided”; in other words, the ‘treatment’ under investigation needs to be individualized and cannot be applied across a group of people without due consideration of individual differences.\textsuperscript{16} Further, the authors point out that knowledge translation is a complex process and in order to change behaviour a comprehensive approach would be needed, within which the provision of evidence is “but one component.”\textsuperscript{17} Finally, the authors admit that because EBM requires lifelong learning, it is possible that “it is not something that can be measured over the short-term.”\textsuperscript{18}

Presumably this is all in contrast to medical treatments. Apparently, according to Straus and colleagues, it is reasonable to suppose that human bodies are so simple, interventions are so straightforward, and outcomes are so unproblematic and uncomplicated that designing research to investigate treatments is no problem at all. Consider, for a moment, the challenges faced by researchers investigating therapeutic counseling for patients suffering from depression. In the design of research, investigators will face challenges in setting outcomes, standardizing treatment, incorporating all elements of the treatment into the analysis and pursuing the counseling for an appropriately long period of time. It would seem that the same sorts of problems raised in research on EBM are encountered in the investigation of some types of medical treatments.

Straus and colleagues suggest that the problems encountered in the evaluation of EBM are idiosyncratic. This is false. The idea that these problems are unique and arise only in the case where we are evaluating EBM is preposterous and even dangerous. While I applaud Straus and colleagues for recognizing and describing the problems encountered in attempts to research EBM, it is astonishing that they do not consider these to be systemic problems worthy of further

\textsuperscript{15} Straus et al., 7.
\textsuperscript{16} Ibid., 7.
\textsuperscript{17} Ibid., 8.
\textsuperscript{18} Ibid. On a personal note, the hypocrisy of these statements reminds me of the way comedian John Stewart uses clips of current U.S. president George W. Bush. He often plays a clip of the president making a dogmatic claim, followed by a more recent clip of him making the opposite claim (just as dogmatically). And then he just pauses and looks confused. I have attempted a more academic tone in my reply, though the sentiment is probably the same.
discussion. Insofar as similar challenges arise in the evaluation of certain medical treatments, those problems will be systematic in evidence-based research. This exposes the hypocrisy of EBM in demanding that all research conform to standards of best evidence, since it is clear that certain types of research questions are not amenable to best standards as defined, and that the results of such research are no less important or meaningful as a result of this difference. Instead of lamenting their inability to perform RCTs, proponents of EBM should get on with the task of investigating EBM with rigorous observational methods, and value the results of such research accordingly. I discuss the importance of the right research design for a particular question further in Chapter Five.

1.2.2. Individualized Care

The roots of medical practice are captured in the famous saying attributed to Hippocrates, “It is more important to know what kind of person has the disease than what kind of disease the person has.” In contrast with this highly personalized philosophy of medicine, EBM suggests that medical practice should be based on the generalized results of clinical research. According to the hierarchy, for instance, meta-analyses of RCTs provide excellent evidence, and case-studies provide the worst. In other words, the larger, more abstract general studies provide better evidence and thus guides to practice than do carefully developed, detailed individual studies. The hierarchy has a clear ‘general to specific’ orientation. As it turns out, however, 1) it is not easy to defend the claim that certain research designs are more generalizable than others 2) this generalizability, if achieved, may actually be a liability in practice.

Let’s start with the standard argument for the generalizability of the highest ranked research methods. Because the RCT is performed on a group of individuals (often quite a large group) and collects average patient data across that group, the average patient data obtained from an RCT is more likely to be generalizable to members of the general population than the data obtained from one particular individual in a case study. This is thought to extend even further in the case of meta-analyses of RCTs. For all we know, the individual in the case study may have

---
19 In the subsequent section of the book, Straus and colleagues consider the limitations of EBM (in about a page) and none of the problems above are even mentioned. Instead, they superficially dismiss a few common concerns about EBM as cookbook medicine, the contention that it is an ‘ivory tower’ approach, and potential economic abuses of EBM. The only limitations on EBM that the authors acknowledge are practical (skill development and time constraints).
been an exception to the norm. In an RCT, the exceptional and the usual are averaged to produce one result: the average efficacy of a particular therapeutic intervention. And in a meta-analysis, the results of a variety of RCTs are averaged again. Proponents of EBM offer the following advice about applying the results of such research to practice,

> [A]sk whether there is some compelling reason why the results should not be applied to the patient. A compelling reason usually won't be found, and most often you can generalize the results to your patient with confidence.

There is a great deal of confidence in the strength of the connection between the results of research and the demands of clinical practice. However, these claims to generalizability are not so straightforward.

Subjects in RCTs are not randomly sampled from the general population. In fact, the subjects enrolled into clinical trials are often quite unreflective of the members of the general population or even the target population for a treatment. While it is sometimes possible to design a trial in which the trial population is purposely matched to the target population for the treatment (at least, as closely as possible), it is uncommon for a number of reasons. First, members of the target population often have a number of comorbidities (concurrent conditions). It is more difficult to isolate a cause-effect relationship when many other variables (including other medications) are added into the equation. These sorts of comorbidities are, accordingly, seen as a liability for the trial. Researchers often explicitly request subjects who have no other conditions and are taking no other medications. Second, members of the target population are often elderly and likely to be put on a treatment for long periods of time or even indefinitely. So the target population would be long-time users of the treatment. However RCTs are usually quite short in duration (ranging from a few weeks to a few months long). There is a gap between the short-term data produced by RCTs and the long-term use by the target population. The data from RCTs, then, is not easily generalized to the standard patient. Third, it is common for researchers to select their study population in a way that allows them to test the healthiest people. The

---

20 In other cases, exceptional cases may dismissed as outliers or attributed to errors in the design of the trial.
21 Guyatt, Sackett, and Cook, “How to Use an Article About Therapy or Prevention.”
22 For instance, people who are mentally ill often have problems with drug addiction and this makes it difficult to isolate new treatments for mental illnesses.
23 These ‘clean’ subjects are increasingly difficult to find in North America. The search for clean research subjects has led many large pharmaceutical companies to the developing world, though while developing world subjects are often free of other medications, they are more likely to have concurrent conditions such as malnutrition.
selection of 18-35 year-old subjects (and sometimes an even narrower range) means that, because in general these subjects are healthier than older people, they are less likely to have adverse reactions to the drug. The healthier a person is, the less pronounced the adverse reaction (in general) and more positive the overall reaction. Again, this means the trial population is different from the target population. Fourth, research trials are often conducted in contexts that differ in significant ways from the contexts of general practice. As a result, therapy may be administered and monitored in a way that cannot be easily replicated. Research is often conducted at facilities with more resources, and implementation of treatments in contexts where resources are scarce is not straightforward. This contributes to a gap between individual patients and the promises of benefit attributed to particular treatments.

Even if we were to set aside these concerns about the generalizability of research from RCTs, we would still be left with concerns about the gap between generalized results and individual patients. This gap between the results of generalized clinical research evidence and complex individual patient care has been raised time and again in medical literature. Feinstein and Horwitz remind us that, “When transferred to clinical medicine from an origin in agricultural research, randomized trials were not intended to answer questions about the treatment of individual patients.”

Patients are idiosyncratic individuals dealing with multiple complex problems. The most scientific medical research by EBM standards produces average generalizations across homogenous populations that, as discussed, often fail to resemble the patient because of inclusion and exclusion criteria. The highest ranked clinical research focuses on large-scale studies designed to determine simple causal relationships between treatments and their effects. In many cases, a particular patient would never have qualified for the study because he or she has comorbidities, or does not meet certain age or gender specifications. As Tonelli points out:

Clinical research, as currently envisioned, must inevitably ignore what may be important, yet non-quantifiable, differences between individuals. Defining medical knowledge solely on the basis of such studies, then, would necessarily eliminate the importance of individual variation from the practice of medicine.

---

If we pay attention to average outcomes, we may lose sight of significant variation within the trial population. It may be that some subjects responded well, while others responded poorly. Though on average this looks like no effect, this is not a full characterization of what has occurred. In fact, important subgroup data will likely be missed. Just as it would be inappropriate for engineers to apply theories in theoretical physics to the construction of bridges without any contextual specifications, it is inappropriate for physicians to apply the results of medical research in practice without attention to the medical context. The applicability of scientific evidence, especially large-scale, single-factor studies, “depends on the individual being conformable to the group in all relevant aspects,” which is rarely the case. On the basis of these concerns, I argue that simply binding medical practice to medical research fails to capture the importance, difficulty, and skills required for re-contextualizing and individualizing medical knowledge. This is not to say that medical practice will not benefit from thoughtful and critical use of a wide variety of results from medical research, but that is not what has been proposed. The critique I am offering here is directed not at a general claim that medical practice should pay attention to medical research (which I take to be a good position), but at the specific rules tying medical practice to medical research in EBM, which I argue fail to capture important distinct elements of medical practice.

1.2.3. Shared Decision-making

The move toward a more inclusive decision-making process that has been occurring in medicine over the last half century has forced physicians to become more aware of the inherent limits and uncertainty of medical decision-making. Recent attention to bioethical principles of autonomy and informed decision-making have pushed the medical model away from doctor-knows-best paternalism and raised new questions about the role of patients in medical decision-making. The more we recognize the role of patients’ values in medical decision-making, the further the original EBM project is from the reality of practice. Even if there are biological similarities between several different cases of a disease, the values and other non-quantifiable factors introduced by each of the individual patients could and should radically shift the medical

---

decision being made. No matter how decisive the evidence is, without patient input there is no clear best decision. Furthermore, the patient’s subjective experience of illness is an important part of the clinical encounter (and the source of his or her desire for treatment) and so should form the basis for discussion about treatment and for the evaluation of objectives and desired outcomes of treatment. These elements of medical decisions require the careful development of listening skills and compassion on the part of the physician and indicate the importance of elements besides scientific evidence to the successful practice of medicine.28

It is unclear how the sort of medical practice advocated by EBM fits with existing standards of shared decision-making and patient-centered care. Shared decision-making implies a shift in the burden of health care work. Patients become part-time (unpaid) health care workers as they search out evidence online, solicit advice from friends, and devote time to careful consideration of their values. Shared decision-making is often justified on the basis of the principle of autonomy (respect for persons). But how are the results of the patient’s work integrated within evidence-based practice? It is clear that patients often highly value anecdotal evidence. If a health problem affected a patient’s friend, sister or grandfather, he or she will be much more likely to attach significant weight to that piece of evidence. Should the anecdotes and intuitions of patients carry any greater weight than those of physicians, which have been explicitly condemned under the EBM model? This seems unlikely, particularly since these anecdotes and intuitions are supremely unscientific by EBM standards. Should the values of patients be permitted to override best research evidence?29 In these situations, it will be tempting for physicians to strongly guide patients to the correct evidence from clinical trials and meta-analyses (indeed, this may be required by EBM). There appears to be, at the very least, a tension between commitments to shared decision-making and the rules of EBM.

The initial formulation of EBM seemed to require an almost algorithmic approach to decision-making according to which individual physicians assess the evidence then apply it to the particular case. The later versions of EBM arguably deal with the need for individualized care, as well as the importance of incorporating patient values by suggesting that medical research should be “conscientiously” and “judiciously” applied to patients. While this tempers

28 The balance between measurement and meaning is an important one for medicine. Upshur, VandenKerkof and Goel.
29 It is worth keeping in mind here that these values may have already been influenced by direct-to-consumer advertising of pharmaceuticals (particularly in the U.S.) and by self-diagnosis on the basis of on-line questionnaires provided by the makers of various medical treatments.
the claim it does not retract it. Medical practice is still assumed to be best when it is based on medical research. The integrative concessions were meant to leave some room in the medical decision-making process for patient values, but patient-centered care and shared decision-making requires more than this. If we take patient-centered care seriously, best evidence will vary depending on the values of the patient and the nature and context of illness. This appears to call into question the basic nature of best evidence. Physicians who emphasize the patient’s role in the decision-making process are likely to resist the standardization implicit in the more scientific approach advocated by EBM.

2. Should Clinical Research Evidence Form the Basis of Medical Decision-making?

According to EBM, clinical decision-making should be based on the results of clinical research. In what follows I consider other possible “bases” for decision-making.

2.1. Not Authority?

In spite of a stated desire to diminish the reliance of physicians on the authority of others, the EBM movement has resulted in the creation of a number of new authorities. The Cochrane Collaboration, founded in 1993, is a volunteer-driven organization whose members perform systematic analyses of the literature and disseminate the results to fellow health care professionals. The main products of the collaboration are systematic reviews – primarily of RCTs – which are updated and added to the Cochrane Database of Systematic Reviews every three months. Various organizations have followed the Cochrane model, such as the National Guideline Clearinghouse in the U.S. and over 40 guideline-producing organizations in Canada, prompting further clarification on the appropriate “guidelines for guidelines” and conferences on methods of systematic review. The production of reviews and guidelines has led to a new type of evidence-based practitioner in recent years. It is not uncommon to find physicians claiming to practice EBM who rely entirely on reviews and guidelines produced by others. EBM users, as

---

30 Canadian Medical Association Infobase.
they have been called, contract out the critical evaluation step originally outlined by EBM proponents.\textsuperscript{31}

This approach to EBM is efficient, but in cutting out the critical evaluation physicians once again put themselves in a position of subservience to authority. In this case it is the authority of those who produce systematic reviews and guidelines, but given that there will always be social and political pressures on those producing the reviews and guidelines, this seems to be a risky endeavor. In establishing pre-digested reviews and guidelines as the new authority, EBM users once again drift away from the demands of critical thinking. Whenever this occurs, there is reason for concern.\textsuperscript{32} These new authorities, draped in claims to scientific objectivity, serve a similar function to that traditionally occupied by older, more-experienced clinicians in the “old paradigm”. The authority of the past has been replaced not by a new democratic spirit in medicine, but by new forms of authority. I will leave analysis of why this has occurred to my colleagues in the social sciences. It will suffice to note that, as EBM becomes the new authority in medicine, it no longer fulfills its own claims to anti-authoritarianism.

\textbf{2.2. Not Pathophysiology?}

EBM proponents make a number of strong statements about the fallibility of pathophysiologic reasoning, for instance:

\begin{quote}
Millions of dollars’ worth of bench research that appeared to show a reduction in atherosclerosis-related oxidative stress with vitamin E therapy was recently tested in an RCT that asked the vital question: Does the vitamin E therapy endorsed by this research really help prevent heart disease? The answer was a resounding ‘No’.\textsuperscript{33}

Predictions about intervention effects on clinically important outcomes based on physiologic experiments usually are right, but occasionally are disastrously wrong. We provide a number of examples of just how wrong predictions based on physiologic rationale can be in Part 2B1, ‘Therapy and Validity, Surprising Results of Randomized Controlled Trials.’\textsuperscript{34}
\end{quote}

\textsuperscript{31} Upshur, “If not evidence, then what? Or does medicine really need a base?”

\textsuperscript{32} A recent article by David Cundiff on the financial interests influencing members of the Cochrane Collaboration makes this point well: Cundiff, 56.

\textsuperscript{33} Sackett, “Time to put the Canadian Institutes of Health Research on trial,” 1414.

\textsuperscript{34} Guyatt, Rennie, 6. Note the implication that clinical trials are never “disastrously wrong”. This will be discussed further in Chapter Four.
Robyn Bluhm has argued convincingly that high level clinical research is dependent on scientific research done in laboratories. Bench science is vital to clinical research because it allows researchers to make causal claims, investigate why outcomes occur and understand subgroup variation in RCTs. This also means bench science is important for clinical practice since on the basis of the advantages just listed it allows physicians to more judiciously apply the average results of epidemiological investigations to individual patients. And this is all on top of the usual claim that bench research is important for hypothesis generation. Pathophysiologic research (bench science research) is lowest on the evidence hierarchy yet it would seem that it could complement research at other levels in the hierarchy in order to provide more robust support for clinical decisions than that provided by research evidence alone. The current evidence hierarchy and admonishments about the use of pathophysiology in clinical practice neglect this relationship.

### 2.2.1. Relationship Between Bench Science and Clinical Research

The uneasy and inconsistent relationship between EBM and the laboratory or bench sciences can be demonstrated in several ways. Common life-saving medical treatments, such as fluids for dehydration, blood transfusions for loss of blood, electric shocks to restart the heart, and antibiotics for pneumonia have never been tested in RCTs. What these interventions have in common is that when they were first attempted there was a dramatic positive effect. Because it is unlikely that unknown confounding factors caused the observed effect, because there is a clear mechanism that seems to underlie the treatment, and because these interventions have been used to great acclaim for decades and in some cases centuries, we do not think it is necessary to investigate them further. As a result, in spite of their impressive pedigree, these interventions have never been rigorously, systematically investigated by researchers in clinical trials. They are currently grandfathered into medical practice under the dubious category of “self-evident” interventions.

---

35 I use the terms laboratory and bench science interchangeably. These sciences (for example, biochemistry and molecular genetics) investigate biological mechanisms.

36 Bluhm, “From Hierarchy to Network: A richer view of evidence for evidence-based medicine.” This relationship goes two ways. For instance, we are especially interested in the particular carcinogens in tobacco smoke because we know from epidemiological research that tobacco smoke can lead to cancer. Vandenbroucke, “Medical journals and the shaping of medical knowledge.”
If EBM proponents are to be consistent in upholding their hierarchy of evidence they have to admit that such interventions are ranked somewhere – probably at the bottom of the evidence hierarchy. The build up of knowledge through trial and error over many years earns only the lowest ranking. This seems strange, given that we are more certain of these treatments than many others offered in hospitals today, and because we think of the hierarchy as indicating what degree of belief we should have about interventions. These interventions are widely accepted on the basis of our current understanding of pathophysiology – an understanding that, it is worth keeping in mind, we didn’t always have. It would, of course, be unethical to randomize people to non-treatment in order to investigate these treatments today. Further, it would be silly, since we know these treatments work. I think even staunch EBM advocates would acknowledge that they know these treatments work. But where is this knowledge coming from? If we are supposed to gain medical knowledge through the hierarchy, and the basic sciences and unsystematic trial and error are ranked near the bottom, does this mean we only have weak knowledge of these interventions? Should we hope for better? This is also problematic when we consider the implications for the future of medicine. Were we to have the good fortune to stumble upon an intervention with the dramatic power of penicillin today, what would we require in terms of evidence of effectiveness? If we require an RCT, are we prepared to accept the loss of lives in the control group? There are undoubtedly going to be lives lost in the initial trial and error evaluation of any treatment, but to follow this with systematic harm through an RCT appears questionable (as it did in the ECMO case outlined in the introduction to the dissertation). These are important questions that get to the heart of the ambiguity of EBM on the value of evidence that we “already know” to be good.

### 2.2.1.1. Homeopathy

The hypocrisy of EBM with respect to bench science is nowhere as evident as in the debates over homeopathy. Though the complexities of the homeopathy debate extend beyond what I will be able to address here, I think it is an important example for bringing out the hidden reliance on bench science within EBM. Homeopathy is an alternative system of medicine that originated in Germany in the late 18th century. It has been experiencing resurgence in North America after a period of relative inattention, which was the result of suppression by the mainstream medical
community in the late nineteenth and early twentieth centuries. Homeopathic practitioners prescribe treatments designed to stimulate the body’s natural healing response by treating “like with like”. The homeopathic substances (parts of plants, minerals or animals) are diluted to infinitesimal amounts through a process of “potentization” and offered to patients, usually in the form of small sugar pills. Remedies first undergo “provings” in healthy subjects and the symptoms that arise in the healthy patients indicate what the remedy will cure in an ill person. So, for example, if diluted rhubarb produces a fever and rash when ingested by a healthy subject, it will stimulate the healing response of someone who suffers from a similar type of fever and rash. There are many other details of this unusual system, but the main point of contention for conventional medical researchers is the ultra-high dilution (UHD) of the ‘active’ ingredients.

Many homeopathic remedies are diluted to the point that not even one molecule of the original substance remains (i.e. beyond Avogadro’s number). Thus the treatment is biochemically nothing more than a sugar pill. Contrary to the dose-response law in chemistry, homeopaths claim that increasing dilution actually makes the remedy more potent. Obviously this does not sit well with chemists. There is also a lack of explanation for how and why homeopathy works, in the trials where it does appear to work. Of course, this lack of explanation is not uncommon in medical research. When coupled with the dose-response problem, though, this gives rise to heated criticisms of homeopathy in the mainstream scientific literature.

What is especially interesting, however, is the mainstream scientific response to RCTs and meta-analyses on homeopathy. While the most recent meta-analysis seems to indicate that homeopathy has no specific effect beyond placebo, there have been earlier meta-analyses – notably one published in *The Lancet* in 1997 – that indicated a slight positive effect of homeopathy. The publication of the 1997 meta-analysis provoked a tremendous response from the medical community. Critics attacked the methods of the review, the credibility of the authors, 

---

37 There are also concerns about the principle that “like cures like.”
38 Ernst and others have suggested that the effects of homeopathy are likely to be non-specific effects. This would be consistent with the position that it is not the biochemical nature of the pill but rather the nature of homeopathic consultations that have an effect (when there is an effect).
39 Some claims about the memory of water or crystalline structures in water have been advanced but have not yet been supported by independent empirical evidence. The most famous of these, which was widely critiqued in the scientific literature, was Michael Schiff’s *The Memory of Water*.
40 The skeptic website Quackwatch has even set up an additional site to specifically address homeopathy: www.homeowatch.org.
41 Linde et al.
and the motivation for the trial. A 2005 meta-analysis of RCTs (also published in *The Lancet*) that came with a press release “homeopathy is no better than placebo” has not received similar critical attention from the mainstream medical community but rather has been widely celebrated and disseminated. The two studies were both meta-analyses: they both used roughly the same methods of evaluation. Yet the only conclusion that was accepted was the one supported by current theories in pharmacology and chemistry. Members of the CAM research community have drawn some attention to this difference in attitude.

Whenever a randomized controlled trial or meta-analysis of RCTs on homeopathy appears in a respected scientific journal, skeptics subject the trial to scrutiny and critical evaluation beyond that which, for instance, a new drug would receive. Singling out homeopathy for this extra-rigorous critique, while refraining from judgement on similarly-designed drug trials cannot be justified by the lack of mechanism for homeopathy since, as I mentioned earlier, it is not uncommon for the full mechanisms of drugs to be unknown (particularly psychiatric drugs).

What is at the core of objections is the fact that homeopathy is inconsistent with current bench science (i.e. laws of chemistry and pharmacology). If the standards of EBM were all that was necessary to evaluate new treatments – that is, if all we really cared about is what works regardless of mechanisms – homeopathy should have been adopted to the same extent as drugs tested in similar trials when the 1997 meta-analysis was published.

I am not suggesting that this should have occurred. What I am saying is that the advocates of EBM can’t have their cake and eat it too, so to speak. They can’t say that only outcomes matter (and devalue research on biological mechanisms) and then turn around and use lack of biological mechanisms, or inconsistency of claims about mechanisms, as the basis for rejecting homeopathy. Pathophysiologic principles do have a role to play in the evaluation of clinical research, and in medical decision-making. There is no reason to suggest that the

---

42 Vandenbroucke, “Homeopathy trials: going nowhere”; Langman; Delblanco.
44 Bell.
45 We often know the mechanism at the basic receptor level (from in vivo studies) and the basic metabolic pathways (from animal research), but we still do not know all of the effects that a drug has in the body, or even the ones that make it effective. Thanks to Robyn Bluhm for clarification on this point.
46 Similarly, they cannot call in Bayesianism to save them by suggesting that homeopathy has low priors while maintaining a commitment to traditional statistics and otherwise ignoring the insights offered by Bayesians on clinical research more generally.
complementary relationship between bench and clinical research warrants the designation of one as the base of medicine.

2.2.1.2. Naïve Empiricism

EBM is only superficially an empiricist approach to medicine. Because it lacks an explicit theoretical framework that connects bench science, epidemiology and everything in between, it is unable to address the problems identified above. In addition, the naïve empiricism of EBM suggests the irony in its choice of moniker. Evidence-based medicine cannot be accused of ignoring evidence. But it fails to get us anything more than that. More specifically, EBM fails to offer us explanations, since explanations require an answer to why questions. But the proponents of EBM seemed aware of this, at least before things spun out of their control. The following quote, taken from a publication by Sackett and colleagues a year before the infamous EBM manifesto says,

If, in the final analysis, the practice of this “science of the art of medicine” is to do more good than harm to patients and clinicians, five additional ingredients must be added to the study of this book. First, its elements must be integrated with those of the other basic sciences, such as morphology, physiology, and biochemistry, as they are applied; were the approaches presented here to constitute the sole scientific basis for clinical action, we would simply be substituting a new tyranny of unachievable methodologic rigor for the old tyranny of unteachable clinical art.

Two things are noteworthy here. First, the authors who wrote this insightful comment describe clinical epidemiology not as an expansion of the existing science of medicine, but as a supplement to the art of medicine. Second, the authors clearly state the importance of understanding clinical epidemiology in context, and yet a year later they initiate a movement that, due in large part to the placement of basic scientific research very low on the evidence hierarchy, led to a derogation of basic scientific research within medicine. The wisdom of this early statement needs to be more widely recognized, and the intervening errors corrected.

---

47 See Chapter One of Robyn Bluhm’s dissertation for further discussion on this point: “The 'hierarchy of evidence' and the structure of medical research: Implications for evidence-based medicine.”

2.2.1.3 Theoretical Framework

The domain of medicine contains three interrelated elements: bench research, clinical research and clinical practice. While EBM has much to say about the relationship between clinical research and practice it has almost nothing – certainly not much that is positive – to say about bench research. As I noted in the last chapter, EBM places basic scientific research at the bottom of the evidence hierarchy, when it is ranked at all. It is also sometimes referred to as a creative realm which supplies hypothesis for proper clinical epidemiological evaluation. The justificatory power of bench science is thought to be very limited. Evidence-based medicine is an attempt to strengthen the connection between clinical research and clinical practice, as well as to specify the appropriate ordering and content of clinical research. The relatively new field of clinical epidemiology, culminating in EBM, prioritized questions that ask whether treatments work. This was done at the expense of significant attention to why and how treatments work. Attempts to answer the latter question, and provide explanations for events, are critical elements in the construction of theories and models in medicine.

What is the cost of having no theory, or of pretending that theoretical frameworks do not exist? First of all, without a theoretical framework (even in the modest sense of having a range of models upon which to draw) it is impossible to respond to the homeopathy problem without hypocrisy. Second, without a theoretical framework it becomes very difficult to identify errors. In effect, what happens in clinical research is continual trial and error, with no feedback mechanism to integrate these findings with other findings, or to generate new hypotheses as a result. This leads to the third cost: without a theoretical framework the chances of innovation are significantly decreased. Because clinical research in this narrow sense is restricted to posing questions about “what works”, it is unable to modify and reformulate the theoretical framework from which these questions should be generated. Explicit attention to the connections between medical theories and clinical research is vital for progress. Marcia Angell, former editor of the New England Journal of Medicine, has argued that the greatest challenge facing medicine today is a lack of innovation. This is not a surprise. This is very much a product of a medical research program that valorizes clinical research at the expense of theoretical research. We cannot simply

---

49 Wennberg.
50 I am indebted to discussions in Paul Thompson and Ross Upshur’s Philosophy of Medicine class at the University of Toronto for clarification on these problems.
repeat a process of trial and error continuously and hope to come to any comprehensive understanding of health and disease.

2.3. Why a Base?

Does medicine really need a base, whether it is clinical research evidence, authority, pathophysiology or something else? Given the number of different elements that factor into clinical decision-making, what is to be gained by prioritizing one? Ross Upshur has argued that the foundationalism assumed by proponents of EBM is ill-suited to the task of guiding medical decision-making.\(^{51}\) Physicians have to incorporate pathophysiologic principles, clinical research evidence, personal experience, intuition, patient values, economic constraints, administrative restrictions, and many other factors when making medical decisions. All of these components are important, and any decision that failed to take all of these factors into account would be poorly informed. The motivation for declaring one of these a “base” appears to be largely rhetorical, given that EBM proponents have more recently admitted the need for integration of clinical research evidence with these other factors. It is also inappropriate, given the hidden reliance on pathophysiology in medical research and the commitment to shared decision-making and respect for patient values, as well as the importance of clinical expertise, in medical practice.

3. Summary

In this chapter I critically evaluated the first two background assumptions of EBM: medicine should be more scientific (given a particular understanding of science) and clinical research should form the basis of medical decision-making. In each case I argued that the assumption is flawed. This does not mean that I am arguing for a less scientific approach to medicine, or that I would direct physicians to ignore the results of clinical research. On the contrary, I think a more nuanced appreciation of scientific reasoning as it occurs in the medical context, and greater awareness of the variety of research methods in medicine would be a vast improvement in medicine. But this is not what is being suggested by EBM, and not what forms the content of the background assumptions presupposed by EBM. Rather, EBM proposes a narrow and restricted

---

\(^{51}\) Upshur, “If not evidence, then what? Or does medicine really need a base?”
account of science, a paternalistic and simplified account of clinical decision-making, and a naïve empiricism regarding the products of medical research. In this chapter I argue that the first two background assumptions of EBM are flawed. In the next chapter I will critically examine the third assumption of EBM.
Chapter Five

Critiquing the Background Assumptions of Evidence-based Medicine: Part II

Proponents of EBM assume that a hierarchy of evidence is needed to guide medical research and practice. However, the particular evidence hierarchy advanced by EBM is only one of many possible hierarchies. If, for example, the values of complexity and individuality were thought most indicative of high quality evidence in medicine, the evidence hierarchy might have been inverted. In fact, recent developments in medicine have led to a proliferation of different evidence hierarchies.¹ In light of the variety of possible and actual evidence hierarchies, the particular version offered by EBM needs to be justified. Advocates of EBM have not been forthcoming on this issue. Because of this, I attempt to reconstruct the most plausible justifications based on comments in the original *JAMA* and *BMJ* articles on EBM, the EBM handbook *Evidence-based Medicine: How to Practice and Teach EBM* and the *Users’ Guides to the Medical Literature.*² According to my analysis, the evidence hierarchy for medical treatments organizes research methods according to their ability to produce results that are: precise, randomized, unbiased, quantified and relatively certain. In this chapter I critically examine each of these justifications for the current evidence hierarchy.³

In the first section of the chapter, I argue that the proponents of EBM ranked research methods in part by their ability to produce precise results. I then critique this assumption. In the second section, I evaluate claims about the epistemic value of randomization and suggest that randomization does not secure many of the epistemic benefits promised by advocates. As a result, randomization is excluded as a justification for the evidence hierarchy. Third, I argue that research methods ranked highest in the hierarchy provide no greater guarantee that biases have been eliminated than those below. This excludes justifications on the basis of objectivity of results. Fourth, I critique the fixation on quantitative methods underlying the evidence hierarchy

¹ Upshur, “Are all evidence-based practices alike?”
³ As noted in the introduction to the dissertation, I am focusing here on the hierarchy for medical treatments. There are also hierarchies for diagnosis and prognosis, for instance.
and demonstrate the value of qualitative methods. Fifth, I consider whether the evidence hierarchy provides physicians with more certainty in practice and argue that uncertainty is ineliminable in medicine and that the evidence hierarchy could enable a dangerous and false sense of certainty. In light of the cumulative effect of these arguments, I conclude that the current evidence hierarchy is unjustified. Finally, I argue that not only is this particular hierarchy unjustified, but that any attempt to rank research methods is incoherent and ultimately detrimental to medical research and practice.

1. Precision

According to the hierarchy outlined in Chapter Three, the highest quality of evidence achievable in medicine is that produced by systematic reviews. Systematic reviews are thought to be advantageous because they minimize bias, assimilate large amounts of information, reduce the delay in translating evidence into practice, and establish the generalizability and consistency of research results. Epistemic advantages, such as the ability to minimize bias, are frequently offered in support of systematic reviews. Systematic reviews are assumed to be minimally biased because the studies they group together are already relatively unbiased: meta-analyses of RCTs are thought to be unbiased because they combine the results of several (already quite unbiased) individual RCTs. Practical advantages, such as the ability to assimilate and translate bodies of evidence into practical guidelines that are ready for use, are also often cited as reasons for preferring systematic reviews. Chiefly, however, quantitative systematic reviews – or meta-analyses – are favoured because they “increase the precision of the overall result” where precision refers to “the measure of the likelihood of random errors (usually depicted as the width

---

4 I am using the version of the evidence hierarchy outlined in Chapter Three. It is the version created by a number of the original proponents of EBM. The arguments in this chapter should apply to most versions of the hierarchy proposed by various international groups, though I do not have the space to properly explore the differences between each of the hierarchies here.

5 Greenhalgh, How to Read a Paper, 116.

6 Straus and colleagues make the following mysterious claim regarding bias in systematic reviews: “We’ve mentioned previously in this chapter the ability of randomized trials to reduce bias. Systematic reviews, by combining all relevant randomized trials, further reduce both bias and random error and thus provide the highest level of evidence currently achievable about the effects of health care. In contrast, systematic reviews of non-randomized trials can compound the problems of individually misleading trials and produce a lower quality of evidence.” Straus et al., 148-9. It is unclear why the explicit methods of systematic review are thought to pool bias in one case while eliminating it in the other. I discuss issues of bias further in the latter parts of this chapter and in Chapter Seven.
of the confidence interval around the result). In what follows I examine the support for this claim. I begin by outlining probabilistic arguments for the precision of meta-analyses. In order to do so I explore the underlying method of hypothesis testing upon which confidence intervals are calculated. Following this I critique the claim that meta-analyses provide better evidence than other research methods. I argue that while meta-analyses do provide results with narrow confidence intervals, there are other meaningful ways of analyzing data that do not make use of confidence intervals. It may worthwhile to surrender some degree of precision for greater attention to the questions that need answering in medical practice. The assumption that precision is among the highest virtues in medical research, though well-meaning, is ultimately misguided.

1.1. Probability

Experimentation yields data. More specifically, RCTs and other highly ranked quantitative research methods produce numerical data which require interpretation and analysis. There are many potentially useful ways of performing this analysis. Of these possible methods (discussed further below), probabilistic methods of analysis are unfailingly preferred in medicine. Further, researchers rely on a very specific subset of probabilistic methods: those that rely on a frequentist account of probability. As medical statistician David Salsburg points out, “It is ‘standard’ to analyze data from a clinical trial using a narrowly defined probabilistic mathematical model.” The very existence of meta-analyses is premised on the assumption that probabilistic models and statistical analysis are the best ways of representing the results of research.

Probability theory is notoriously controversial. This controversy occurs right down to the foundational definitions of probability offered by various camps. There are many such conflicting definitions; I will focus on two of particular interest: frequentism and subjectivism (or Bayesianism). Frequentists propose a relationship of identity between the probability of an event and its frequency. Under this account probability just is frequency: the number of events divided by the number of possible outcomes. Frequentist probability specifies the number of times an event will occur in the long run of nearly identical experiments. Subjective accounts of

---

7 Greenhalgh, *How to Read a Paper*, 116, 123.
8 Salsburg, “The Use of Statistical Methods in the Analysis of Clinical Studies,” 17.
probability suggest that probability just is degrees of belief. Agents assign probabilities according to their subjective creedal state. Of these two accounts of probability, the frequentist definition “lies at the heart of the standard hypothesis testing formulation. Unless one accepts the frequentist interpretation, the reasons for treating the standard formulation as an optimal procedure disappear.” I turn now to an explanation of hypothesis testing in order to establish the connection between frequentist probability and corresponding types of statistical analysis and hypothesis testing.

1.1.1. Hypothesis Testing

Despite the variety of possible analytic tools, in the medical context the standard is frequentist statistical analysis, relying on frequentist accounts of probability. In the reporting of results, medical researchers often focus on discussions about p-values and, more recently, confidence intervals. This focus comes about as a result of reliance on what is known as “hypothesis testing”. According to this approach, data are analyzed with respect to a statistical null hypothesis, which posits that there will be no effect of the treatment. The null hypothesis is a useful fiction, because when employed it creates the appearance of testing hypotheses against each other, since the real hypothesis (that the treatment is effective) can be compared to the null hypothesis (that the treatment has no effect). The worry with hypothesis testing is that we’ll either accept a null hypothesis that is actually false (type two error) or reject a null hypothesis that is actually true (type one error). Type two errors will eventually get corrected as further research is performed. Type one errors lead researchers to think that treatments are effective when they are not. This type of error is of particular concern in medicine, where treatments proven effective are quickly manufactured and prescribed to patients. Since we cannot completely rule out this type of error we are forced to choose a “level of significance”, which indicates what the chances are of rejecting a true hypothesis. Significance is how likely a value is to be due to chance, so we want it to be as low as possible. We determine the level of significance on conventional (non-statistical, non-empirical) grounds. This significance level is called the p-value and it is generally set at 0.05. What this means is that if you were to do 100 tests, you would reject a true hypothesis in 5 of those tests. We have no way of knowing when

---

9 Ibid., 21.
this has occurred. Significance tests (for example, the t-test, the chi-squared test, and the Wilcoxon Rank Sum test) are only possible on data from randomized trials.

P-values are a measure of the weight of evidence against the null hypothesis, and if the p-value is small enough (below a set threshold), the null hypothesis is rejected. The limitation of hypothesis testing that emphasizes p-values is that it provides only yes or no answers. (Is treatment x more effective in this group of subjects than no treatment? Yes/No.) It does not provide information about magnitude of effect, fluctuations over time, subgroups with interestingly different responses from the norm, and so on. This is problematic: “The implication of hypothesis testing – that there can always be a simple ‘yes’ or ‘no’ answer as the fundamental result from a medical study – is clearly false and used in this way hypothesis testing is of limited value.” Further, such tests “often do more damage than good” because they gloss over the complexities of clinical practice in favour of simple answers. In response to these concerns, there has been greater attention paid to confidence intervals in recent years.

Confidence intervals give an estimated range of values within which the trial result is likely to fall, based on a certain set confidence level (usually 95%). The shift from p-values to confidence intervals (“probabilities of probabilities”) is an improvement because confidence intervals, unlike p-values, allow researchers to include effect size (or magnitude of effect) in their assessment of research results. Confidence intervals indicate a range of values within which the correct result likely lies, and so show us the range of possible effect sizes that are compatible with the results of research. Understanding the magnitude of the effect is vital for assessing clinical significance, rather than mere statistical significance. Confidence intervals are thought to correct for this false dichotomization of results by indicating a range of values (rather than one p-value) that is more reflective of the generalizability of results to the population. Confidence

---

10 Quite a lot has been written about the failings of significance levels, or p-values. P-values are said to: fail to capture what matters about data, fail to represent evidence because they are dependent on only one hypothesis (and evidence requires information about the alternative hypotheses), depend on a fallibilistic theory of scientific inference that is in tension with commitments to positive induction in science more generally, fail to address effect size (unless supplemented by confidence intervals), obscure the data through requirements that investigators monitor their hidden actions and intentions during the trial, limit us to thinking that our beliefs can only be weakened by new data (rather than strengthened), and finally give inconsistent results in cases where the evidence should be the same and consistent results when the evidence should be different. As a result, journals such as *Epidemiology* have taken a principled stance against their use. In spite of criticism from a variety of perspectives over almost ninety years, p-values continue to feature prominently in almost all of the medical literature produced today. Studies in top journals will almost always cite p-values, though usually in conjunction with confidence intervals. Gardner and Altman, 746; Goodman and Royall, “Evidence and Scientific Research”; Rothman, “Writing for Epidemiology”; for examples of recent studies citing p-values see Pierce et al.; Hildesheim et al. in the latest issue of *JAMA*.

intervals still depend on the sort of approach characterized here as hypothesis testing, however, and though the results are less rigidly constrained to yes/no declarations, adding ranges of values around the dichotomous yes/no answer is still regarded as somewhat restrictive.

### 1.2. Meta-analyses and EBM

Meta-analyses reduce random errors. Recall that a meta-analysis is “a statistical synthesis of the numerical results of several trials which all addressed the same question.”\(^\text{12}\) Claims that meta-analyses are “precise” refer to narrow confidence intervals. The prominent placement of meta-analyses at the top of the hierarchy indicates the importance of narrow confidence intervals within EBM since confidence intervals indicate the precision of results. Recall that the hierarchy consistently places meta-analyses of RCTs above individual RCTs, meta-analyses of cohort studies above individual cohort studies, and meta-analyses of case-control studies over individual case-control studies. This indicates the value placed on narrow confidence intervals even when lower-quality studies are being analyzed. The evidence hierarchy assigns the highest level of status to meta-analyses because meta-analyses make use of statistical tools of analysis that permit the creation of large sample sizes and thus the minimization of random error. In the world of EBM, meta-analyses are king.

The official logo of the international Cochrane Collaboration is a pictorial representation of the results of a meta-analysis of seven RCTs. The illustration is a circle with a vertical line down the middle and a series of seven horizontal lines at different intervals from the top to the bottom of the circle. The vertical line – the “line of no effect” – indicates the probability that the two treatments being compared (usually treatment and placebo) have the same effect. This is a standard method of representing confidence intervals and the cumulative results of meta-analyses. In this case it culminates in a diamond to the left of the vertical line, near the bottom of the circle. The individual RCTs in this meta-analysis investigated a short course of corticosteroids to prevent premature labour. What is interesting about the meta-analysis that forms the Cochrane Collaboration logo is that while only two of the individual RCTs produced confidence intervals that excluded the possibility that the treatment had no effect, the cumulative data from all seven trials indicated a clear positive effect of the treatment. (The cumulative result

is indicated by the diamond.) According to the meta-analysis, the treatment reduces the odds of neonatal mortality by 30-50%. This result is more confident and more positive than any of the individual RCTs because when the results of all seven trials are brought together the meta-analysis can draw on this cumulative sample size. The pooled large sample size means that researchers performing a meta-analysis can secure narrow confidence intervals for their results. For this reason, meta-analyses of RCTs are highly regarded within the EBM hierarchy. The Cochrane Collaboration logo highlights the strength of meta-analyses in securing precise results.

The tendency to prefer analyses that bring together a body of literature is not only evident in the evidence hierarchy we are currently examining. EBM advocates have argued that even meta-analyses are too inaccessible to the busy practitioner: “What busy practitioner has time to use evidence-based resources if the evidence is presented in its original form or even as detailed systematic reviews?” Straus and colleagues suggest a “4S” approach. Naturally, it is hierarchically organized:

```
Systems
Synopses
Syntheses
Studies
```

At the lowest level of this hierarchy we have the individual studies (RCTs, cohort studies, case-control studies, and so on – with no mention of the lowest ranked evidence such as case studies and anecdotal evidence, which appear to have dropped off the bottom). Above that are syntheses such as meta-analyses of RCTs (of the sort produced by the Cochrane Collaboration). Even higher than meta-analyses are synopses. These are the abstracts of meta-analyses found in evidence-based journals. And if the abstract is just too much to handle, Straus and colleagues helpfully point out that “in some circumstances, the title [of a review] provides enough information.” Their interest is in conveying the results of research – not the methods or messy details – as effectively as possible. Finally, at the top of the hierarchy we find computerized decision support systems, which would ideally integrate the lower levels of evidence in one location. Computer systems are not yet able to do this, but according to Straus and colleagues this is would be an ideal approach.

---

13 Cochrane Collaboration Website: www.cochrane.org
14 Straus et al., 37. Emphasis added.
15 Ibid.
It may be the case that the abstraction from the data of original research seen in this 4S hierarchy is motivated largely by issues of expediency and practicality. In order to be justified as a good route to knowledge, however, the 4S approach ought to at least protect the production of knowledge (if not enable it). As mentioned above, the diminished possibility of random error in meta-analyses is thought to provide this epistemic justification. In what follows I will argue that there are reasons to question this focus on precision – particularly when we are sacrificing the relevance of research.

1.3. Critique

Statistician David Salsburg has argued that the narrow focus on hypothesis testing in medicine misses out on a whole range of other possible methods of analysis available within statistics. Medical statistics, it is argued, “reflects activity in only a small portion of the statistical world.” He calls on medical researchers to “cast off the awkward cloak of hypothesis testing and treat the data as an exercise in estimation of parameters and the identification of reasonable subsets of patients.” One consequence of this shift, he argues, will be a greater relevance of trials for clinical practice. He cites the overuse of a limited range of statistical concepts as part of the reason for the slow uptake of research in clinical practice.

In order to make this point clear, Salsburg presents examples of the sorts of questions asked by physicians in practice:

2. If a patient is going to respond to treatment A, how long will it take for the response to manifest itself and what can be monitored to know whether such a response has occurred?

3. What patient characteristics are there that will identify patients most likely to respond and patients most likely to suffer adverse reactions? (This assumes, of course, that there are some patients who will respond, so it might be appropriate to apply a preliminary hypothesis test before chasing down will-o’-the-wisps via regression analyses.)

4. If an adverse reaction is going to occur, what are its early manifestations, and what is the general pattern of the hazard function (increasing, decreasing, or constant over time)?

---

17 Ibid.
18 Ibid.
Standard analysis does not yield answers to these sorts of pressing questions, which Salsburg identifies as most valuable in practice. Instead analysis yields answers on the efficacy of particular treatments relative to the null hypothesis. Salsburg argues, “It is of little use to society to have a simple yes or no to the question of whether a new treatment is ‘dangerous’ or whether a new treatment ‘works’.” What is needed, instead, is “to use these studies to identify subsets of patients with specific response patterns and to estimate degrees of effect and the time course of events.” These sorts of results would be considerably more relevant to the practical questions identified above. And yet these are precisely the opposite of the sorts of meta-analyses deemed best evidence by the hierarchy of EBM and captured in the 4S approach.

Salsburg argues that reliance on a narrow frequentist approach to statistical analysis leads researchers to pose irrelevant questions (i.e. those unlike the three just identified). The data produced by a trial should be analyzed in such a way that it answers the questions physicians need answered in order to practice medicine. Yet because of assumptions about the importance of a certain specific frequentist statistical analysis – to the exclusion of other types of analysis – the data is not usually analyzed to provide these sorts of answers. Salsburg rightly concludes: “When what is essential to a particular type of mathematical modeling appears inappropriate to the experimenters who are interested in scientific conclusions, there are serious philosophical problems at hand.” Lest he be accused of speaking too theoretically, he proposes a few practical alternative methods of data analysis. For instance, we can isolate those subjects who respond to treatment and trace their baseline characteristics to see if they have anything in common. If they do, this might help us identify a subgroup of subjects who are most responsive to the particular treatment. Alternatively, we might measure the way the disease pattern fluctuates over time in the control group and model this typical pattern. Then we would be able to compare the subjects in the treatment group against the typical pattern. Neither of these methods, nor any other non-probabilistic method of analysis, is currently popular in medicine. The probabilistic model is dominant because we want broad-based application of the results of research, yet as pointed out in Chapter Four, even the probabilistic results produced in meta-analyses cannot be applied

---

19 Ibid.
20 Ibid.
21 Ibid. 18.
22 Ibid.
23 Ibid.
24 Ibid.
broadly to the general population. If we aren’t actually gaining the ability to apply the results broadly, why is there a continued emphasis on these sorts of results?

Consider the following case. A new treatment is tested in an RCT. Analysis of the results indicates a p-value of 0.049. Researchers conclude that the treatment is effective for the pre-specified outcome. The confidence interval is narrow because the sample size was large, so the results are thought to have been well-established, and physicians are advised to prescribe the treatment to the general population. Closer attention to the results, however, yields interesting further data. For instance, while the treatment group did better overall, within the group half the members did slightly worse, while the other half did quite a lot better. On average this meant that the group did somewhat better (enough to be statistically significant). Surely it is important to understand the subgroup differences in this trial. Imagine, for instance, those in the treatment group who responded very well were men and the others who did not were women. This data can too quickly be ignored, and even lost in meta-analyses that analyze results of multiple trials in terms of average p-values and confidence intervals. Physicians prescribing the treatment, however, would be better served by closer attention to these differences. In this example, the relevant subgroup data will be analyzed probabilistically so this re-analysis still relies on probabilistic analysis – though applied to different data. A more innovative re-analysis would consider non-probabilistic methods of interpreting the results as well. This is precisely what medical statisticians such as Salsburg are arguing. There are better ways of getting answers to important clinical questions than to unreflectively analyze data from RCTs searching for the “rare mice” of low p-values or narrow confidence intervals.

1.4. Statistical Style of Reasoning

The choice of probabilistic model underling EBM is undefended and I would suggest that the choice has more to do with trendiness and educational inertia than the provision of sound arguments. It is worth briefly attending to the ways in which popular use of a theory or method can shape the development of an entire field. Ian Hacking has suggested that there are a variety of styles of reasoning, including (but not limited to): simple mathematical postulation, experimental exploration and measurement, analogy, comparison and taxonomy, historical
Styles of reasoning, according to Hacking, are created in order to meet the changing needs, interests and goals of various communities and are often, as he puts it, “defended by bluster and insidious patience.” As a style of reasoning evolves within a particular community it gradually becomes fixed as an objective route to knowledge. That is, it is assumed without question that the reasoning style is a route to knowledge:

[I]t needs no support or rhetoric, for as it assumes self-confidence it generates its own standard of objectivity and its own ideology. It starts by being pushed and shaped by social vectors of every sort; we end with a self-sustaining mode of knowledge. It becomes less something molded by interests and more an unquestioned resource upon which any interest must draw, if it ever hopes for the accolade of objectivity. And it further determines how people conceive of themselves and their world, opening new horizons, but also constraining the possible forms of knowledge.

Hacking goes on to argue that statistical reasoning is self-authenticating in precisely the way just outlined. We believe that the results of statistical calculation are, on the whole, accurate. But our measure of accuracy in this case is, of necessity, also determined using statistical principles. We have no true number independent of any statistical method against which we could measure, to use a famous example, the population of New York. We determine the reliability of the statistical style of reasoning using further statistics (in the medical context confidence intervals perform this function). Hacking argues that, though it appears circular, it is the same apparent circularity found in all styles of reasoning and would be better described as boot-strapping. The case of statistical reasoning just makes this boot-strapping obvious, since it is so transparent.

One of the lessons I draw from Hacking’s analysis, which overlaps with the concerns raised by Salsburg, is that we need to be aware of this process of self-authentication. We ought to be aware of the ways in which the process tends to exclude alternative approaches, even when, from the outside, we can see that those alternatives might be valuable. This analysis points out an important implication of this understanding of the self-authentication of styles of reasoning: the constraint of possible forms of knowledge to those that fit the accepted style. If, in medicine, meta-analyses are highly regarded – to the point that physicians are actually counseled to look only at meta-

\[25\] Hacking, “Statistical Language, Statistical Truth and Statistical Reason,” 132. This particular list was originally proposed by Crombie in 1978.
\[26\] Ibid.
\[27\] Ibid.
\[28\] Ibid., 144.
analyses and not bother with original research – and if this assumption rests on a characterization of styles of reasoning that is unnecessarily narrow, we have reason to call this assumption into question.

1.5. Conclusions on Precision

Salsburg instructively writes, “many branches of scientific research flourish without probabilistic modeling or with, at most, a minimal use of averages and standard deviations. These include molecular biology, mathematical biology, physiology and physical chemistry.” 29 To the extent this is the case it is because other formal techniques have been developed that take their place. Why then the assumption that best evidence in medicine has a small p-value and narrow confidence interval? While it may turn out to be the case that (frequentist) probabilistic methods of analysis are very useful in medicine, this does not mean that other methods of analysis are to be ignored. There are other methods of statistical analysis that draw on Bayesian accounts of probability. More to the point, though, there are entirely non-probabilistic methods of analysis available. Hacking identifies several other ‘styles of reasoning’ and Salsburg points out the types of questions and corresponding methods of analysis that are obscured by the attention to a certain narrow range of probabilistic methods in medicine. Attention to the value of other methods of analysis may yield answers to more pressing clinical questions than those traditionally answered in meta-analyses of RCTs.

2. Randomization

One of the principal divisions in the evidence hierarchy is that between randomized and non-randomized trials. 30 If a trial is randomized, it is ranked near the top of the hierarchy. If not, it is

30 Discussions of randomization in clinical research refer to the random allocation of subjects to some number of treatment and control groups (usually one of each). They do not refer to the selection of a random sample of the general population for the study (in fact, clinical trials are often notoriously unrepresentative of the general population because of inclusion and exclusion criteria). Random sampling is arguably related to our ability to generalize from the results obtained on the particular subjects under investigation to the population as a whole, while random allocation is thought to balance treatment and control groups and isolate a cause and effect relationship between treatment and outcome. The following discussion addresses random allocation. Random allocation is achieved by tossing a die, flipping a coin, drawing a card from a deck, or through more complicated measures such
below. So what epistemic benefits are attached to randomization? The value placed on randomization is most evident in the sharp line drawn between the RCT and the lower-ranked cohort study. Given that cohort studies are also controlled trials (they have treatment and control groups), can be double-blinded (though this depends on the type of intervention, as it does for RCTs), can be analyzed under the intention-to-treat protocol, and have an identical causal inferential structure (eliminative induction), the only feature distinctive of RCTs is the random allocation of participants to the two groups. In what follows I will consider four possible arguments for randomization, and outline various critiques that have been raised against these arguments. I argue that randomization does not secure the epistemic benefits it is thought to secure. In conjunction with the other arguments made in this chapter, I argue that the RCT should not have a position of privilege in the evidence hierarchy. First, I will establish the EBM position on randomization.

2.1. Randomized Trials and EBM

Randomized trials are consistently placed at the top of various versions of the evidence hierarchy. The following quotes are representative of the attitude toward randomization by proponents of EBM:

[W]e owe it to our patients to minimize our application of useless and harmful therapy by basing our treatments, wherever possible, on the results of proper randomized controlled trials.\(^{31}\)

To ensure that, at least on your first pass, you identify only the highest quality studies, you include the methodological term ‘randomized controlled trial (PT)’ (PT stands for publication type).\(^{32}\)

If the study wasn’t randomized, we’d suggest that you stop reading it and go on to the next article in your search. (Note: We can begin to rapidly critically appraise articles by scanning the abstract to determine if the study is randomized; if it isn’t, we can bin it.) Only if you can’t find any randomized trials should you go back to it.\(^{33}\)

---

\(^{31}\) Sackett et al., *Clinical Epidemiology*, 195.

\(^{32}\) Guyatt, Sackett, and Cook, “How to Use an Article About Therapy or Prevention,” 59.

\(^{33}\) Straus et al., *Evidence-based Medicine: How to Practice and Teach EBM*, 118.
The first two statements were made by prominent members of the EBM working group, and the chatty advice proffered in the third appears in the 2005 edition of the official EBM handbook. Claims that EBM does not privilege RCTs and that advocates do not tell physicians to ignore other sources of evidence are contradicted by these recent statements. In addition, careful examination of the guidelines used in the evaluation of research evidence indicates a persistent tendency to set aside all studies that are not RCTs, despite claims to the contrary.\textsuperscript{34}

\subsection*{2.2. Non-randomized Trials Exaggerate Treatment Effects}

A number of arguments are made in favour of randomization. The argument that non-randomized trials overestimate treatment effects is pervasive in the medical literature. Sackett and colleagues write, “It is when asking questions about therapy that we should try to avoid the non-experimental approaches, since these routinely lead to false positive conclusions about efficacy.”\textsuperscript{35} It does not take anything more than elementary logic to point out the circularity in this argument. In order to establish that RCTs are more accurate than other methods the authors start by assuming that RCTs are more accurate than other methods, show how the results of non-randomized trials differ from the results obtained through RCTs and conclude that RCTs are better. Data indicating a persistent difference of this sort (were it to exist), however, would be consistent with the claim that non-randomized trials are producing more accurate accounts of events and that randomized trials systematically underestimate treatment effects. The fact that there is some apparent difference does not immediately lend itself to interpretation one way or the other. These sorts of arguments beg the question.

Furthermore, when offering empirical evidence for the claim that non-randomized trials produce different results on average from randomized trials – a claim that can be made without circularity – Sackett and colleagues provide very little empirical evidence. Recent investigations have indicated that there is no significant difference between the results of well-conducted RCTs, cohort studies and case-control studies. John Concato, Nirav Shah and Ralph Horwitz examined the results of different trial designs investigating the same clinical interventions and found that,

\textsuperscript{34} Grossman and Mackenzie exposed this problem using the example of the MERGE guidelines in their 2005 paper.

For the five clinical topics and 99 reports evaluated, the average results of the observational studies were remarkably similar to those of the randomized, controlled trials…The results of well-designed observational studies (with either a cohort or a case-control design) do not systematically overestimate the magnitude of the effects of treatment as compared with those in randomized, controlled trials on the same topic.36

In fact, they discovered that there was less variability in the results of non-randomized trials on the same intervention than within the set of randomized trials. It is not the case that non-randomized trials exaggerate treatment effects.

2.3. Non-randomized Trials Are Misleading

This claim is sometimes formulated as part of the last argument, though it has a broader scope,

Because the randomized trial, and especially the systematic review of several randomized trials, is so much more likely to inform us and so much less likely to mislead us, it has become the ‘gold standard’ for judging whether a treatment does more good than harm.37

What quite often happens is that advocates of EBM provide a few examples of non-randomized trials or pathophysiological reasoning that we now know to have been flawed.38 With the benefits of hindsight we now know that interferon for cancer, hydrocortisone for acute myocardial infarction, and freezing of the stomach for duodenal ulcers are not effective treatments. Once a few of these examples have been listed, proponents make statements like the following: “The literature is replete with examples where a new treatment was highly touted by its proponents and enjoyed a good deal of popularity for a time, but did not withstand the test of a well-controlled randomized clinical trial.”39 As Ross Upshur puts it, however, “that a hypothesis turns out to be wrong is nothing new nor [anything] to be particularly ashamed about. It is how science works.”40 Moreover, there are many recent examples of treatments that were approved in RCTs

36 Concato, Shah and Horwitz, “Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs,” 1887. See also Benson and Hartz, “A Comparison of Observational Studies and Randomized Controlled Trials.”
38 We saw some of these arguments against pathophysiology in the Chapter Four.
39 Grage and Zelen, “The Controlled Randomized Trial in the Evaluation of Cancer Treatment – the Dilemma and Alternative Designs,” 37. In addition, EBM advocates often compare poorly conducted non-randomized trials to well conducted RCTs in order to make the benefits of RCTs more apparent.
40 Upshur, “Thin Warrant.”
and which we later found to be harmful through long-term observational studies. Cox-2 inhibitors (for instance, Vioxx®) and spironolactone are two examples of treatments that produced positive initial results for specific conditions in RCTs but which we have since realized are harmful to some patients. In the case of Vioxx®, millions of patients were affected by the medical community’s failure, among other things, to believe the results of observational studies which indicated increased risks associated with its use. Were we to follow the reasoning of EBM proponents, we might be tempted to suggest that randomized trials routinely mislead us. What we should be reminded of instead is the uncertainty and fallibility of all medical research, regardless of the choice of research method. As the trial by Concato and colleagues demonstrated, on average the results of well-designed randomized and non-randomized trials were the same.

2.4. Non-randomized Trials Cannot Balance Groups

It is often asserted that only randomization can balance unknown confounding factors in treatment and control groups. Confounding factors are variables that affect the dependent variable under investigation in a trial. In order to understand this claim, consider first the importance of control groups. If we want to know whether a new flu vaccine is effective we will not get much information by giving it to one group of people and recording the results. This is because without a comparison group we have no sense of whether, for instance, a 10% rate of the flu in the vaccine recipients is more or less than what it would have been otherwise. In order to know whether the flu vaccine is effective we need to compare rates of subsequent infection in people who did and did not receive the shot. Hence the need for a control group. And not just any control group will do. It would obviously be problematic if we assigned elderly people to the

---

41 Upshur, “Looking for Rules in a World of Exceptions.”
42 Concato, Shah, and Horwitz.
43 To use a popular example, if murder rates are highly correlated to ice-cream consumption in a particular population, this does not necessarily mean that ice-cream consumption causes murders, or that murders drive people to consume ice-cream (though this could turn out to be the case). Rather, correlation of these factors may indicate that the two are both caused by another variable – say, hot weather. If we design a study to investigate the relationship between ice-cream consumption and murder, and fail to control for the state of the weather, we might end up making inappropriate claims about the causal relationship between murder and ice-cream consumption. In this example, weather conditions are a confounding factor.
44 In spite of this well-known fact, flu vaccines in Ontario have been offered free to everyone without any systematic attempt to track and compare results across populations. Ross Upshur, personal communication.
control group and young people to the treatment group (or vice versa) since we have reason to think that factors such as age influence the odds of contracting the flu. Treatment and control groups that are imbalanced with respect to these sorts of ‘known’ (suspected) confounders make it difficult for researchers to isolate the cause and effect under investigation. It leaves open the possibility that the confounding factor, not the treatment, is producing a particular effect.

We can control for known confounding factors by deliberate matching. For instance, we can make sure that there is an equal distribution of elderly and young people in our two groups. However, it is impossible to know that we have matched for all confounding factors because there are an infinite number of possible confounders – many of which, no doubt, we will not have any reason to suspect. It may turn out to be the case that blonde-haired people get the flu more often than other members of the general public. If the treatment and control groups are imbalanced with respect to blonde-haired subjects we have the same problem we started with. As long as we only make use of deliberate matching, it will always be possible for the groups to be imbalanced on an unknown confounding factor. If so, the success or failure of the treatment may be a result of this imbalance and not due to the properties of the treatment under investigation.

Advocates of randomization are quick to step in at this point; after all, it is only randomization of study participants that can control for all known and unknown confounding factors. This claim goes back to Fisher, who writes that the significance test can be “guaranteed against corruption” by the use of randomization. As Worrall points out, this is far too strong a claim, and Fisher and other statisticians who have made similar claims must have been aware of this. The two groups can, at best, be balanced for all factors only “in some probabilistic sense.” The defensible weaker position tempers its claims with statements like “as balanced as possible” and refers to the “tendency” for balance rather than any guarantee. More specifically, Worrall argues, randomizers are arguing that it is improbable that the two groups are imbalanced with respect to any one particular unknown confounder. However,

Even if there is only a small probability that an individual factor is unbalanced, given that there are indefinitely many possible confounding factors, then it would seem to follow

---

47 Ibid.
that the probability that there is some factor on which the two groups are unbalanced…might for all anyone knows be high.\textsuperscript{48}

In order to begin to address this problem (though it can never be fully overcome), the randomization would have to be repeated an indefinite number of times. In RCTs, randomization is done once. Given that I am studying the standards of evidence used in the medical context, I am interested in claims made about RCTs as they are used in research, not in claims made about hypothetical RCTs in ideal worlds. As a result, then, randomization does not create the conditions for justified reasoning by eliminative induction: “The premise that the experimental groups were \textit{probably} balanced does not imply that the differences that arise in the clinical trial were \textit{probably} due to the experimental treatment.”\textsuperscript{49} If the two groups are only probably balanced (and it is clear that this is the best we can hope for), it is no longer possible to claim that we are reasoning by eliminative induction, because we have not eliminated the possible options, but only made them less likely. In an RCT, there are two options: the causal hypothesis which claims that the treatment has an effect (whether positive or negative), and the null hypothesis which claims that there is no effect. Once the null hypothesis has been rejected, that is, once there is some significant result, the researcher knows that some causal factor is at work. In order to be able to claim that the treatment under investigation is that causal factor, all alternatives have to have been eliminated. As it turns out, though, the alternatives are never eliminated.\textsuperscript{50}

To sum up, the standard position is that non-randomized trials may be unbalanced with respect to unknown confounding factors and so produce skewed results. Only randomized trials are said to ensure that unknown confounding factors are distributed randomly among the two groups. But it turns out that the groups likely are not balanced, and in fact the odds are very good that, given an indefinite number of possible confounders, any given RCT will be imbalanced on at least one factor.\textsuperscript{51} Running the trial over again (and again) would help with this problem, but is impractical. Meta-analyses might be thought to save the day here, but in fact they pool data from similar trials asking similar questions – not replicated trials. Randomization does not balance

\textsuperscript{48} Ibid., S324. Worrall draws on work by Dennis Lindley here.

\textsuperscript{49} Howson and Urbach, \textit{Scientific Reasoning}, 197. Emphasis added.

\textsuperscript{50} This will not come as a surprise to anyone familiar with the problem of induction. In fact, the common claims about the ability of RCTs to secure causes betray a lack of appreciation for the problem of induction. While paralysis in light of the problem is unhelpful, so is pretending that you have secured facts when you have not.

\textsuperscript{51} Worrall, “\textit{What evidence in evidence-based medicine?” S324.}
treatment and control groups with respect to all possible confounding factors.\textsuperscript{52} This does not mean that randomization is entirely ineffective – it still makes it \textit{less likely} that confounding factors are at play. This has some epistemic value. But this value is much more limited than generally recognized, and does not provide a basis for ranking randomized methods categorically above carefully matched or historically-controlled trials.

\textbf{2.4.1. Aside: Table 1}

This brings us to one of the odd features of RCTs as they appear in medical journals today.\textsuperscript{53} RCTs are consistently published with a Table 1, which is meant to indicate whether randomization has been successful or not. In the table, researchers select possible known confounding variables and see whether they were, in fact, evenly distributed into the two groups as a result of randomization. While groups that are balanced on suspected factors may be unbalanced on other factors, “we are reassured when the known prognostic factors are well balanced.”\textsuperscript{54} This betrays a significant misunderstanding of randomization. The mistake is captured perfectly in a Dilbert cartoon:

\begin{quote}
Accounting Troll 1: Over here we have our random number generator.
Troll 2: Nine nine nine nine nine nine nine.
Dilbert: Are you sure that’s random?
Troll 1: That’s the problem with randomness: you never can be sure.\textsuperscript{55}
\end{quote}

A sequence of nines may be just as random as something that appears stereotypically random (i.e. has no discernable pattern). You cannot assess whether randomization has been successful or not. Table 1 also betrays a lack of faith in the powers of randomization. After all, if it works, it works. Since you cannot go back and make sure it worked with respect to the unknown confounders, what is the point of checking on known variables (rather than matching for them in the first place)? It is a pointless exercise. If this is driven by (justified) lingering concerns about

\textsuperscript{52} It is important to note that there are statistical techniques beyond matching which can adjust for confounding factors in non-randomized trials; stratification, multivariable adjustment, propensity scores, and instrumental variables are some of the most common. So randomization isn’t necessarily the best we can do, either.

\textsuperscript{53} Thanks to Ross Upshur for bringing this problem to my attention.

\textsuperscript{54} Guyat and Rennie, \textit{Users’ Guide to the Medical Literature}, 62.

\textsuperscript{55} Adams.
randomization in general, then there is a need for explicit recognition of this and a better solution.

2.5. Non-randomized Trials Identify Correlations (Not Causes)

Only randomized trials are thought to be capable of establishing genuinely causal relationships between treatments and effects; studies lower on the hierarchy get at “mere correlation”. Causation is a complex concept; as such, it is important to be clear on what is meant by a cause in this context. Two types of causes are common in medicine: mechanistic causes and probabilistic causes. Mechanistic causes are provided by bench research in biochemistry, genetics, physiology and other basic sciences and are thought to be especially stable because they hold in all cases (not just selected sub-populations, however carefully or randomly selected). Probabilistic causes establish strength of association between dependent and independent variables in a given population, ideally in repeated studies.

In Chapter Four I outlined the EBM position on pathophysiologic mechanisms and bench science. While mechanistic and probabilistic causes might intuitively be thought of as complementary ways of understanding the empirical world, the evidence hierarchy identifies probabilistic causes as epistemically superior. Claims about the special causal abilities of RCTs refer to probabilistic causes and ignore the possibility that mechanistic causes could be just as well-established, just as epistemically strong, and just as useful in medical practice. Consider Bradford-Hill’s nine criteria for causation: strength of association, temporality, consistency, theoretical plausibility, coherence, specificity, dose-response relationship, experimental evidence, analogy. Of these criteria, several explicitly relate to mechanisms: temporality, theoretical plausibility, coherence and experimental evidence all rely on a characterization of a cause as a mechanism of some sort. Many of the remaining criteria relate to probabilistic causes. It is unclear why some of these criteria (those that are probabilistic) have been elevated

---

56 Russo and Williamson. There is considerable debate over the ontological status and unity/disunity of the concept of causation. Russo and Williamson, for instance, argue that while there are two distinct types of evidence in medicine there is only one unified account of causation. A comprehensive catalogue of causes is offered in Parascandola and Weed. There may also be other types of causes relevant to the health context. I do not offer a comprehensive overview of causation in this dissertation as it is beyond the scope of this project.

57 Russo and Williamson.

58 Hill.

59 Russo and Williamson.
within EBM while others (those that are mechanistic) have not. In addition to the neglect of other types of causes, the assumption that RCTs uniquely isolate probabilistic causes runs into its own problems.

Claims about the causal powers of RCTs begin with the assumption that randomization controls for all known and unknown confounding factors. We have already seen why this is problematic: because there are indefinitely many possible confounding factors it is probable that, in any particular trial, the two groups are, in fact, unbalanced. Following this, defenders of the special causal ability of RCTs make claims about the epistemic powers of actual RCTs based on what would happen in ideal RCTs. The presence of the phrase “in the long run” betrays the slide to theoretical claims that have no particular bearing on RCTs as they are done in the medical context. If we were to randomize forever, the limiting-average effect of the treatment would yield information of the sort desired by RCT enthusiasts. However, RCTs are only ever done once. Even on the infrequent occasion when an RCT is repeated it is done on different subjects, in a different context. It is not, strictly speaking, replicated. Unfortunately, “there is no reason to think that any actual randomized trial gives the same results as would be got from the ‘limiting-average’.” Because of the number of variables at play it is more likely that, were the trial to be run many times, each set of results would be slightly different. So while we might be justified in making claims about the causal powers of randomization in the long run, in the short run (which is all we have) those powers are nonexistent. It is not just that it is logically possible for RCTs to fail to establish causation (we already knew that based on the number of conflicting RCTs, after all). It is that we never know how close they have come to doing so. This is not significantly different from the sorts of claims that can be made about the results of, for instance, well-conducted historical trials. There is no special access to causes granted only to RCTs.

Claims that RCTs isolate causes, while other methods identify merely correlations, have undefined and undefended accounts of causation that unfairly denigrate mechanistic causes, depend on arguments about the ability of randomization to balance groups on known and unknown factors (which we’ve already identified as problematic), and rely on characterizations of ideal RCTs (such as the indefinite repetition of the trial) that are never attainable in practice. All research methods that make use of probabilistic methods of analysis have some ability to get

---

60 “[R]andomization will ensure that all other causes of R are probabilistically independent of T in the long run.” Papineau, D. “The Virtues of Randomization,” 446-7. Emphasis added.
at probabilistic causes and it may be the case that, in cases where an RCT is the best method for a particular question, it is especially good at narrowing down the possible causes, but this does not mean that RCTs have a unique capacity to identify causal relationships.

2.5.1. Aside: Side-effects

Extensive confusion about causation in the medical literature is reflected in claims about the real effects and side-effects (or adverse effects) of treatments. Clinical trials claim to demonstrate that a treatment works by causing a particular (desirable) outcome. When this has been provisionally established, they tack on a list of possible side-effects that are only stripped of a more honorific title of primary effect in virtue of not having been selected as the endpoint of interest when the trial was designed.62 As a result, researchers publish results purporting to demonstrate the efficacy of a particular treatment, even though the rate of certain side-effects is as high (or higher!) than the desired outcome.63 If a treatment has multiple effects, there is something deceptive in labeling one of them the primary effect and others side-effects. This is not just a theoretical problem. Treatments are being prescribed for one condition and causing a handful of others. This is compounded when doctors prescribe further treatments to address the symptoms and side-effects of the first treatment. The rate of deaths caused by adverse drug reactions is alarming, even with the most conservative estimates and the most robust contextualization, including recognition that we have no other treatments for many diseases.64 Confusion about causation contributes to this problem. RCTs are designed to investigate single cause-effect

---

62 When trials are designed, the standard view has it that we can only draw reliable conclusions about outcomes we determine in advance. This prevents data dredging.

63 For example, in a 2005 article on low-dose aspirin for the primary prevention of cardiovascular disease in women over 45 (published in the prestigious New England Journal of Medicine) the authors emphasize the statistically significant benefits of aspirin in the prevention of strokes. A closer examination of the original data indicates that there were approximately four times more adverse events (gastrointestinal bleeds of any sort) than beneficial events with respect to strokes. And yet the analysis selects only a small subset of the gastrointestinal bleeds (those resulting in blood transfusion) as their adverse event rate. Even with this smaller number, the benefits and risks come about even – certainly there is no clear benefit. In spite of all of this, the emphasis in the conclusion of the article is on the potential benefits of low-dose aspirin in the prevention of strokes, though the authors do acknowledge the importance of individualizing decisions for particular patients. Ridker et al.

64 The most widely cited figure on deaths due to adverse drug reactions is a conservative estimate of 100,000 per year in the United States. Including hospitalizations and injuries without death pushes this number over 2 million. This would translate to about 10,000 death per year in Canada, or 200,000 events overall, assuming the two countries are roughly comparable on this issue. Many scholars regard this as a very low estimate and suggest that the number is more likely several times larger. Lazarou, Pomeranz, and Corey, “Incidence of adverse drug reactions in hospitalized patients.”
relationships between treatments and desired outcomes. Some people read the results of such studies without contextualizing them and assume that all that happens is that a specific treatment causes a desired effect. The causal story is much more complicated. Of course, conscientious physicians are aware of this, but it is interesting how the way the research is presented tends to play into such fantasies about single causes and effects.

2.6. Conclusions on Randomization

A number of claims have been made about the epistemic benefits of randomization. In the evidence hierarchy for medical treatments, randomized trials are consistently ranked above non-randomized trials. It should not be inferred, on the basis of the critique of RCTs just outlined, that RCTs are bad trial designs or should not be pursued. What I and others argue is that RCTs are not necessarily better than other methods. Insofar as the evidence hierarchy elevates randomized methods above all other methods categorically, the hierarchy is flawed. This will not mean the end of RCTs, but rather the end of the claim that they are superior to all other methods merely in virtue of the fact that they are randomized. It will allow for the results of well-conducted cohort or case-control studies (and the results of qualitative studies and case studies) to outweigh the findings of RCTs in particular cases. As Howson and Urbach put it, “removing the absolute requirement for randomization is a significant step which lifts some severe and, in our view, undesirable limitations on acceptable trials.” Sometimes non-randomized trials are better than randomized trials. Randomization does not confer the advantages it is often claimed to confer upon evidence. Thus, randomization does not justify placing one trial above another in the evidence hierarchy.

3. Managing Bias

One of the justifications of the evidence hierarchy is that it ranks research methodologies according to their ability to eliminate bias. In what follows I will investigate whether evidence ranked highest in the hierarchy is, by design, less biased than that below. I argue that it is not. In statistical terminology bias is, “a systematic distortion of an expected statistical result due to a

Howson and Urbach, Scientific Reasoning, 202.
factor not allowed for in its derivation; also, a tendency to produce such distortion.” The narrow concern of statistical methodology is with confounding factors as possible sources of bias. I have already discussed confounders in the section on randomization above. In this section I will consider bias in a broader sense. According to the Oxford English Dictionary, bias is: “an inclination, leaning, tendency, bent; a preponderating disposition or propensity; predisposition towards; predilection; prejudice.” This broader definition includes what I have been referring to as social or contextual values in previous chapters. I will begin with a description of the claims about bias made by proponents of EBM.

3.1. Bias and EBM

The evidence hierarchy rests on the notion “that it is possible to rank methods of inquiry by their susceptibility to bias.” The EBM working group writes about the systematic attempts to record observations in an “unbiased” fashion as one of the key features distinguishing clinical research from clinical practice. According to the Canadian Task Force on Preventive Health Care, the evidence hierarchy is designed to “place greatest weight on the features of study design and analysis that tend to eliminate or minimize biased results.” Of all of the available methods, the RCT is thought to be least subject to bias. I turn now to a description of the range of biases that impact clinical research.

3.2. Catalogues of Bias

In 1979, David Sackett wrote an article entitled “Bias in Analytic Research”, in which he offered a preliminary catalogue of the sources of bias that distort medical research. The paper was written in response to the growth of case-control studies in the medical literature at the time. One of Sackett’s aims was to encourage further methodological developments in order to improve the design of case-control studies. Another of his aims was to demonstrate that cohort studies are

66 Oxford English Dictionary Online.
67 Ibid.
69 Evidence Based Medicine Working Group, 2421.
70 Canadian Task Force on Preventive Health Care.
71 Sackett, “Bias in Analytic Research.”
preferable to case-control studies, when possible, because their methodology allows for greater prevention and control of bias. This was based on an understanding of the profound (negative) impact biases could have on the results of a trial.\textsuperscript{72} The original catalogue lists 56 biases that arise over seven stages of research.\textsuperscript{73} Sackett provided detailed empirical evidence of a sample of 9 biases that occur during the course of the trial, along with suggestions for preventing and minimizing these forms of bias. Sackett acknowledged that many of the other sources of bias are difficult to address. In fact, he only offered mechanisms for dealing only with 9 of the 35 biases of sampling and measurement (and these represent only two of the seven stages of research), and even those with only partial success.

A recent second edition of a well-known users guide to Randomized Controlled Trials provides a more contemporary catalogue of the types of biases that can influence medical research. The authors acknowledge that there are potentially infinite sources of bias, and outline over 60 of the most common types, at five stages of research:

1) Prior to the trial (e.g. choice of question bias)
2) During the trial (e.g. outcome choice bias)
3) Reporting the trial (e.g. selective reporting bias)
4) Disseminating the trial (e.g. publication bias)
5) Uptake of the trial (e.g. reader bias)\textsuperscript{74}

It is worth keeping in mind that Sackett’s original catalogue dealt with biases affecting case-control studies (and, for comparison, cohort studies). One might expect that there would be fewer biases influencing RCTs, since they are thought to be an improvement over the cohort design. But that would underestimate the creativity of the human mind, both in coming up with new research loopholes and in identifying and cataloguing them after the fact.

\textbf{3.3. Bias in RCTs}

\textsuperscript{72} Sackett highlights the problem of bias in an attempt to encourage the improvement of the methodological standards of case-control studies. In a prescient statement, he writes “The failure to respond here may lead to the publication of a rash of ill-conceived, seriously flawed case-control studies and a subsequent rejection of the entire approach by an inflamed scientific community.” It is ironic that Sackett proudly led this ‘inflamed’ community largely to reject case-control studies a decade later (they ended up quite low on the evidence hierarchy).
\textsuperscript{73} The seven stages of research in which bias can occur (as identified by Sackett) are: “(1) In reading-up on the field, (2) In specifying and selecting the study sample, (3) In executing the experimental manoeuver (or exposure), (4) In measuring exposures and outcomes, (5) In analyzing the data, (6) In interpreting the analysis, and (7) In publishing the results [and back to (1)].” Ibid., 61.
\textsuperscript{74} Jadad and Enkin, \textit{Randomized Controlled Trials}. 
As I discussed earlier, the (causal) inferential structure of the RCT is almost identical to the cohort study, though the cohort study is consistently ranked below the RCT in various versions of the EBM hierarchy. The superiority of RCTs is often illustrated with reference to two forms of bias: selection bias and ascertainment bias. The RCT is said to be the only method that can control for these forms of bias. This is thought to partly justify the higher status of RCTs in the evidence hierarchy. In what follows I will take a close look at the biases that RCTs alone are thought to be able to address.

The authoritative CONSORT statement defines selection bias as: “systematic error in creating intervention groups, such that they differ with respect to prognosis. That is, the groups differ in measured or unmeasured baseline characteristics because of the way participants were selected or assigned.”\(^7\)\(^5\) Jadad and Enkin suggest that this is the one bias RCTs can truly claim to control better than other trial designs. Even philosophers who are otherwise very critical of RCTs are willing to concede that they have some special epistemic powers with respect to selection bias.\(^7\)\(^6\) This form of bias can occur when selecting subjects for the trial from the general public.

In the early days of clinical research, before randomization was popularized, medical researchers attempted to achieve balanced treatment and control groups by alternating the allocation of patients to the two groups as they were enrolled into the trial. The problem with this was that physicians modified their behavior depending on whether the next patient was to be enrolled into one group or the other. Physicians would, on occasion, refrain from inviting patients into a trial when they knew the next participant would receive placebo or purposely enroll patients who were more or less likely to do well on the treatment into one or the other group depending on what they hoped to establish with the results of the trial. In order to deal with selection bias of this sort, it is important (at least, in these types of trials) that researchers institute some form of allocation concealment. By concealing the allocation of enrolled patients from the physicians doing the intake for the trial, this form of bias can be minimized – at least insofar as it differentially affects the groups (physicians may still reject subjects for fear they will imbalance the overall results, or even design the trial so that it explicitly excludes people

\(^7\)\(^5\) CONSORT Statement.
\(^7\)\(^6\) Urbach; Worrall “Why There’s No Cause to Randomize.”
unlikely to benefit from the treatment in order to make it more likely that the results are positive overall).

Allocation concealment is secured independently of randomization. In fact, a trial can be randomized and yet fail to have allocation concealment (this was of great concern to the proponents of EBM, who pushed for explicit statements about allocation concealment in published studies). If the physician conducting the trial is aware of the randomization sequence, then there is no allocation concealment. And a non-randomized cohort study can have concealed allocation (it is just a matter of keeping the allocation criteria – whatever they may be – from the physicians doing the intake). Perhaps more to the point, other research designs such as case studies and qualitative research (in-depth interviews, for instance) do not encounter this bias. When allocation concealment is achieved, it is not because the allocation was randomized, but rather because it was successfully concealed. RCTs are not unique in controlling for selection bias.

How about ascertainment bias, then? It is defined as the “systematic distortion of the results of a randomized trial as a result of knowledge of the group assignment by the person assessing [the] outcome, whether an investigator or the participant themselves.” It is hard to imagine why this would be uniquely achieved in RCTs. Ascertainment bias arises in the patient reports and analyses of the trial as it nears completion. If either the patient or physician is aware of the group the patient ended up in, this may lead them to report more positive, or more negative, results. For instance, a patient may overstate his or her improvement in order to gain praise from the physician, or the physician may ask fewer questions or adopt a more detached attitude in order to get more subdued reports from patients in the placebo group. As with all biases, these may be conscious or unconscious. The mechanism for addressing such bias is blinding. It is important to ensure that the participants, physicians, and even analysts are blind to the allocation of treatment and control groups. Double and triple-blinding are mechanisms for diminishing the effect of ascertainment bias. And blinding is not unique to RCTs, nor even always possible in RCTs. RCTs on lifestyle and surgical interventions, for instance, are often not

---

77 They may encounter a similar bias, but because they do not have control groups, or even groups at all in some cases, the bias is not exactly the same. Qualitative researchers also have a number of ways of dealing with the selection of patients to studies. Furthermore, none of the strategies for dealing with selection bias in quantitative research are perfect, and this includes allocation concealment. Researchers have been known to break into locked desks, hold envelopes up to the light and steam open envelopes in order to get at allocation patterns.

78 CONSORT Statement.
blinded. Even if we were to imagine that blinding is unique to RCTs, this would not necessarily indicate that RCTs are more objective, since other research methodologies, such as historical or retrospective cohort or case-control studies are not faced with this potential bias (though they encounter others).

Let us be charitable for a moment and suppose that RCTs do manage to control for one or two types of bias. Let’s say my arguments about selection and ascertainment bias are refuted, or that other biases are found that only RCTs can manage. This does not mean the collapse of my argument that RCTs are no better at controlling for bias than other study designs. This is because even if RCTs did manage to control for one or two biases that no other trial could address, it is likely that the same could be argued for other research methods. Perhaps there are unique forms of bias faced only by case studies, which only case studies can address. This does not mean that case studies are more objective or value-free. In a sense, what such studies have done is create a problem and then solve it. But other trials may not be faced with that problem. So it is not meaningful to suggest that such trials are more objective simply because they have conquered one or two particular biases.

Furthermore, it is not clear to me why we should think that all biases are created equal. Some biases appear to have a more significant distorting affect on the results of experiments than others. Even if RCTs could control for selection bias in a way that no other trial could, and all trials faced concerns about selection bias, it may be that selection bias is a small player in the realm of bias. I would argue, along these lines, that the forms of bias earliest in the design of research are likely to have the greatest impact on the objectivity of results – and these are the forms of bias that we have no specific rules or mechanisms in place to address. Choices of study topic, research method, and population of interest have a tremendous effect on the results of research, but are relatively subjective.

The catalogues of bias produced by Sackett, Jadad and Enkin serve to remind EBM advocates of the precarious and imperfect nature of all evidence, including that produced by the gold standard method in medical research. There are dozens of possible biases influencing research at any given moment. While we might manage to control for some of these sources of bias through strict methodological requirements of research, there are many others that will remain. This provides a timely reminder of the need for critical reflection on the results of even
the best research, and the importance of educating physicians and patients on how to deal with uncertainty. It also encourages us to take a closer look at what we mean by bias in medicine.

For every bias, or negative value, on the list of biases provided by Enkin and Jadad, there is a corresponding positive value. We avoid hidden agenda bias because we assign a positive value to open agendas. We avoid publication bias because we assign a positive value to equality or justice in the evaluation of publications. These positive values, in turn, are justified on the basis of epistemological assumptions about how to best arrive at knowledge in the scientific domain. As Longino puts it, “[T]he question of whether social values can play a positive role in the sciences is really the wrong question. Social and contextual values do play a role, and whether it is positive or negative depends on our orientation to the particular values in question.” This indicates a need for greater attention to the role of values in medical research. Identifying and evaluating biases that have a negative impact on inquiry is an important project, as is the re-education of health care professionals regarding the positive and productive role of values in inquiry. Without an appreciation for this range of roles, the job of weeding out negative values will be superficial and ultimately unhelpful. Thus, there is a need for transparency about all values in research. In line with the CCE norms, we need to recognize that these pervasive values and assumptions need to be critically discussed and evaluated in order to ensure that idiosyncratic assumptions and values are not unduly shaping research.

### 3.4. Conclusions on Bias

Even in the most methodologically rigorous studies, significant biases can occur. Especially in the case of larger studies and meta-analyses, “While trials can be scrutinized for gross errors of methodology, and excluded if such are found, less obvious but still material differences in procedure may not be apparent to a scrutiny diffused over a number of protocols.” These differences in procedure can be manipulated to achieve different effects. Researchers have been quite inventive at coming up with new ways to subvert legitimate inquiry (without committing outright fraud) including: sub-optimal dosing of the competitor’s drug in a head-to-head trial, testing on young people (in order to minimize the number of adverse reactions), publication of

---

80 Black, 1.
only positive results, publication of only part of the results of a trial, analysis on the basis of secondary endpoints when primary endpoints do not indicate a significant effect of the treatment, and so on.\textsuperscript{81} Specific drug studies were examined in one review, and the authors found that, “while there was no relationship between drug company funding and study quality, those studies which were supported by a drug company were significantly more likely to have an outcome favouring the drug of interest (98% vs. 79%).”\textsuperscript{82} In other words, despite equally good methods in the different studies, bias still played a role in the research outcome. This is no surprise given the catalogue of biases in RCTs outlined by Jadad and Enkin.

Even if we were to set aside global social concerns about political and economic influences on the direction of research, and the individual biases introduced by researchers, the catalogues of specific biases identified by Sackett, Enkin and Jadad suggests that bias can and does play a pervasive role in research. These findings have direct implications for an evidence hierarchy that claims to diminish bias through methodological rigour. RCTs are widely thought to be more objective (less confounded) than other trial designs. What is particularly interesting is that the one or two biases on this list that RCTs are thought to be especially designed to eliminate are either equally well managed by other methodologies (because they are unrelated to randomization) or they are not encountered by other methodologies. As such, the claim that RCTs are, by design, more objective than other trials is false.\textsuperscript{83} And bias is still a problem in need of a solution.\textsuperscript{84}

4. Quantifiability

Noted EBM critic, Alvan Feinstein, has written about the “distraction of quantitative methods” in medicine, which has been particularly common in the last three decades.\textsuperscript{85} Quantitative

\textsuperscript{81} Parker, “Whither our art?” 274; Angell, \textit{The Truth About the Drug Companies}; Sackett and Oxman. A case could be made that some of these choices constitute fraud, which no methodology can prevent. However, because many of these choices have to be made in the course of research-as-usual, it may be difficult to make the case that one choice is fraud while another is not (for instance, one choice of inclusion criteria is inappropriate while another is not). There are grey areas here between honest research and fraud, and that is precisely where these manipulations come into play.
\textsuperscript{82} Norman, “Examining the assumptions of Evidence-based Medicine,” 141.
\textsuperscript{83} Grossman and MacKenzie, “The Randomised Controlled Trial: Gold Standard, Or Merely Standard?”
\textsuperscript{84} Mechanisms for dealing with biases in research will be discussed in greater detail in the final chapter of this dissertation.
\textsuperscript{85} Feinstein, “Clinical Judgment Revisited: The Distraction of Quantitative Methods.”
outcomes, research methods and methods of data analysis are celebrated within the evidence hierarchy. This focus on quantification in medicine relates to the previous discussion on bias, since rules and numbers can perform similar functions in inquiry. According to Theodore Porter, one of these functions is the projection of value-freedom. Decisions made “by the numbers”, or by explicit rules, appear to be fair and impartial.\textsuperscript{86} Evidence at the top of the hierarchy is both produced by explicit rules designed to eliminate bias (discussed above) and quantified. In what follows I will explain the EBM position on qualitative research, provide examples of the many uses of qualitative research in medicine, and conclude with thoughts about why qualitative research has been devalued within EBM. If it is true that “[a]n analytical, quantitative and critical approach to medical data is on the rise,” I investigate whether this focus on quantitative research is justified.\textsuperscript{87}

4.1. Qualitative Research and EBM

The original evidence hierarchies either made no mention of qualitative research or ranked it near or at the bottom alongside anecdotal evidence and case-studies.\textsuperscript{88} The most recent handbook on EBM devotes “a few words” and less than three pages to the assessment of qualitative research, noting in particular that qualitative research “may provide us with some guidance in deciding whether we can apply the findings from quantitative studies to our patients.”\textsuperscript{89} This is hardly an enthusiastic endorsement. It likely reflects a more common perception of qualitative research as “frivolous, faddish, and devoid of real substance, value and utility.”\textsuperscript{90} Similarly, in the original “Users’ Guides to the Medical Literature” series in \textit{JAMA}, Mita Giacomini and Deborah Cook (on behalf of the EBM Working Group) suggest that “qualitative studies offer an alternative when insight into the research is not well established or when conventional theories seem inadequate.”\textsuperscript{91} Again, the endorsement here is weak: qualitative research can be helpful only when other (presumably better) methods have failed. Recently there have been efforts to create separate qualitative evidence hierarchies and corresponding guidelines. The Campbell

\textsuperscript{86} Porter, \textit{Trust in Numbers}, 8.
\textsuperscript{87} Schattner and Fletcher, “Research Evidence and the Individual Patient,” 1
\textsuperscript{88} Gray, \textit{Evidence-based Healthcare}.
\textsuperscript{89} Strauss, S.E. et al., 143.
\textsuperscript{90} Sandelowski, “‘To Be Of Use’,” 125.
\textsuperscript{91} Giacomini and Cook, “Users’ Guides to the Medical Literature: XXIII. Qualitative Research in Health Care A. Are the Results of the Study Valid?” 357.
Collaboration – which mirrors the quantitatively focused Cochrane Collaboration – was established in 2000 to analyze the results of qualitative research. In spite of the advances in the visibility of qualitative studies, qualitative researchers face an uphill battle for recognition and qualitative research is far less likely than quantitative research to influence clinical practice.92

4.2. What is Qualitative Research?

Like quantitative research, qualitative research sets out to describe, understand and explain empirical phenomena. At the core of all qualitative projects is an emphasis on context. Where quantitative research asks questions about frequency and quantity, qualitative research asks ‘what?’, ‘why?’ and ‘how?’ While quantitative inquiry is often concerned with hypothesis testing, qualitative inquiry is more often interested hypothesis generation or theory construction. Qualitative inquiry is characterized by certain research frameworks (e.g. constructivism), types of data (e.g. stories), research methods (e.g. grounded theory), techniques for data collection (e.g. interviewing), techniques for data analysis (e.g. constant comparison), and techniques for interpretation (e.g. hermeneutics).93 Qualitative research methods not only generate data using unique methods such as focus groups, they also study naturally occurring events using ethnography or video recordings, and provide theoretical analyses of documents, texts and practices.94 Qualitative research is frequently described as “deficient” when compared to quantitative research. Because qualitative research draws upon methods and processes seen as contaminated by quantitative researchers (such as disciplined subjectivity) it is often characterized as “limited”.95

In contrast with these negative characterizations, qualitative research has many uses in the medical context. First, qualitative research can provide critical analysis of key medical concepts, and so shape their use in research and practice. The redefinition of “non-compliance”

---

92 There are a few notable exceptions within the EBM community. I will discuss those below.
93 Sandelowski, “‘To Be Of Use’,” 125. Sandelowski goes on to point out that the separation between quantitative and qualitative research is not nearly as simplistic as it appears, though it is often useful to use the simple categories in discussions. In reality, however, “ ‘Qualitative’ researchers count, use statistical methods, and qualitatively test hypotheses. ‘Quantitative’ researchers engage in flights of fancy, seek the perspectives of their subjects, and interpret verbal data,” 126.
94 Barbour, “The role of qualitative research in broadening the ‘evidence base’ for clinical practice.”
95 Sandelowski notes that while the methods of purposeful sampling are referred to as ‘non-probability’ sampling, you would be hard-pressed to find reference to probability sampling as ‘nonpurposeful’. The value judgement implicit in such descriptions reflects a general distrust of qualitative methods in the more-quantitative literature, 126.
as “self-regulation” is one example. This sort of research helps explain apparently illogical patient behaviour (such as refusal of the flu shot) in different, more patient-centered terms. The goal of qualitative research is “not the mere accumulation of information, but rather the transformation of understanding.” Because changes in understanding often reform rather than build upon existing knowledge, it can seem like qualitative research does not advance with the same sort of momentum as quantitative research, but this misses out on the important conceptual work being done by qualitative research, and the sorts of diverse perspectives highlighted by qualitative methods.

Second, qualitative research can provide complementary analyses that contextualize quantitative results. For instance, qualitative studies can provide explanations for the unusual or unexpected results of quantitative trials, such as the reasons for poor compliance among a particular subgroup in a clinical trial. Additionally, qualitative research can suggest mechanisms for outcomes discovered in observational research. For instance, one group of qualitative researchers investigated the reasons why childhood accidents differed among certain subgroups and suggested a possible explanation: social class differences in the interpretation and implementation of safety rules. The epidemiological data alone indicated only that there was a difference, not why it existed. The additional information also informs possible solutions and makes them more likely to succeed. The understanding of social processes demonstrated by qualitative studies is critical to the implementation of health policy. Qualitative research can also generate hypotheses for further quantitative investigation.

Third, qualitative research can stand on its own as a useful guide to clinical decision-making. For instance, qualitative research that identifies non-evidential factors influencing clinical decision-making (such as apparently irrational desires and beliefs) has helped physicians understand what it is that patients are seeking when they come in to a doctor’s office. One study, which drew on semi-structured interviews with patients and practitioners, as well as audiotapes and interviewer notes, identified 14 types of misunderstandings related to patient information. All of the misunderstandings were associated with potential or actual adverse

---

96 Conrad, “The meaning of medications.” Conrad argues that assumptions about ‘deviance’ implied in the assessment of a patient as noncompliant can be recast, from the patient’s perspective, as a positive description of autonomous and informed decision-making about their personal care.
97 Sandelowski, “‘To Be Of Use’,” 125.
98 Sparks, Craven, and Worth, “Understanding differences between high and low childhood accident rates areas.”
99 Britten et al., “Misunderstandings in prescribing decisions in general practice.”
outcomes. These included cases where information known to the patient was not passed along to the physician, such as relevant facts about the patient’s medical history or desires not to receive a prescription for further medication. Furthermore, information about treatment side-effects, dosage, and alternative options was not communicated to patients by physicians. And the list of miscommunications goes on. The researchers also propose solutions. They suggest that many of the miscommunications were associated with the patient’s lack of participation in the consultation, and they propose further qualitative work into appropriate educational interventions.100

Another qualitative study, which investigated the processes leading to inappropriate use of pharmaceutical drugs in elderly inpatients, found that inappropriate prescribing, counseling and other interventions were most common in cases where the physician had a passive attitude to learning, when review of treatment was driven by acute considerations, and when a paternalistic relationship existed between the doctor and patient.101 These findings were determined on the basis of semi-structured interviews with doctors, nurses and pharmacists, and observation of inpatients on the ward, as well as focus groups with selected inpatients. These sorts of findings are vital for a proper assessment of patient needs, and could prevent the over-prescription of medications (which is a serious problem) and improve awareness of the value of shared decision-making and more egalitarian physician-patient relationships.

Finally, qualitative research is the only kind of research that is capable of providing answers to certain types of questions. For example, the following research questions would be best answered with various qualitative methods: “What are the attitudes and behaviors of ICU [intensive care unit] clinicians toward clinical practice guidelines?”; “What is it like for patients leaving the ICU?”; and “How do critical care unit staff decide which patients are admitted to the ICU?”102 A recent study, which used grounded theory and self-administered questionnaires, examined the experiences of family members of patients in the ICU. The satisfaction of family members regarding the communications they had with physicians reflected preferences for three types of clinicians’ statements: “assurances that the patient will not be abandoned before death; assurances that the patient will be comfortable and will not suffer; and support for [the] family’s

100 Ibid.
101 Spinewine et al., “Appropriateness of use of medicines in elderly inpatients.”
102 Sinuff, Cook and Giacomini, 106.
decision to withdraw or not withdraw life support.” Satisfaction was even greater when these types of comments were repeated many times by physicians. The information gleaned from this mixed-methods study is much more helpful than anything possible using quantitative methods alone.

The results of qualitative research, when designed to address questions of clinical importance, are unquestionably relevant to medical decision-making. Because these results are directly relevant to clinical decision-making, and provide support for different aspects of decision-making, such as effective communication, they also demonstrate the role that qualitative research can play in bridging the research-practice gap. It is odd that this potential role was neglected by designers of EBM. Insofar as the questions asked in the studies just mentioned could only have been investigated through qualitative methods, the evidence hierarchy may prevent such questions from being asked or answered.

4.2.1. Aside: Why Focus on Quantitative Research?

Why did the proponents of EBM de-value or even ignore qualitative methods? Most obviously, the movement arose out of the clinical epidemiology movement, which because of its roots in epidemiological methods had a traditional focus on quantitative measurements. But there are reasons beyond this. Theodore Porter has written extensively on the use of numbers and methods of quantification in the sciences. Porter argues that numbers have always been used to indicate rigor and universality and to assist researchers in gaining the trust of others. Communication between researchers is made possible by the use of numbers because they are regarded as impartial and because they are accessible across divisions of language and culture. But while these somewhat more defensible reasons for quantification are present, they are usually accompanied by less transparent motives. Porter writes,

While…numbers and systems of quantification can be very powerful, the drive to supplant personal judgment by quantitative rules reflects weakness and vulnerability. I interpret it as a response to conditions of distrust attending the absence of a secure and autonomous community.  

--

103 Ibid., 109.
104 Porter, Trust in Numbers, xi.
Numbers can hide all sorts of questionable assumptions and practices. The boundaries of medicine have been contested by various outsiders over the years, from midwives to naturopaths. This puts mainstream medical professionals constantly on the defence. Quantifying data and making medicine appear more scientific are mechanisms for protecting it from “quacks”. Scientific communities that choose to elevate methods of quantification are adapting in response to challenges from within and outside their community. I will be discussing these challenges, particularly with respect to the rise of complementary and alternative medicine, in the next chapter. The focus on quantitative methods is not necessarily problematic, however, unless it has epistemic costs. However, those costs, as it turns out, can be high; the trade-off for quantification is often accuracy:

There is a strong incentive to prefer precise and standardized measures to highly accurate ones. For most purposes, accuracy is meaningless if the same operations and measurements cannot be performed at other sites. This is especially true, and especially urgent, where the results of research are to be put to work outside the scientific community.\(^{105}\)

Porter goes on to say that results are “put to work” in the field of medicine more so than any other field. So while accuracy \textit{should} be highly valued in medicine (one would think), it may be given up for highly quantified results. These concerns dovetail with those raised at the beginning of the last chapter over the gap between research and practice in medicine. Accuracy on questions that matter may be sacrificed for generalizations on questions that do not matter as much. Similarly, accuracy on important questions about meaning and care may be sacrificed for quantitative data on outcomes. Pursuing this path fails to acknowledge Alvan Feinstein’s insight, “not everything that is measured is important and not everything that is important can be measured.”\(^{106}\)

\textbf{4.3. Clarification}

There have occasionally been articles, or comments within articles, written by members of the original EBM working group, in which the authors have indicated the need for a more inclusive

\(^{105}\) Ibid., 29.
\(^{106}\) Feinstein. Quoted in Daly, 47.
attitude toward qualitative research than that represented by the EBM handbook or other guides to EBM, or than is represented in the quotations above. In particular, David Sackett has offered a more reasoned position on the matter. Sackett and Wennberg write, “Our thesis is short: the question being asked determines the appropriate research architecture, strategy, and tactics to be used – not tradition, authority, experts, paradigms, or schools of thought.”\textsuperscript{107} As an example they write that a question about patient preferences in choice of treatment would be best investigated with in-depth interviews. They even make the eminently reasonable statement, “Each method should flourish, because each has features that overcome the limitations of the others when confronted with questions they cannot reliably answer.”\textsuperscript{108} I am obviously sympathetic with this position; in fact, I think it is entirely correct. The question then becomes: is this position representative of EBM? Sackett is often thought of as the father of EBM, and certainly until 2000 he was one of the movement’s most charismatic proponents. Why not take his statements as reflective of the core of EBM?

The view on qualitative research that Sackett presents, though laudable, is not representative of the core of the EBM position for two reasons. First, Sackett’s position is inconsistent with the authoritative guides produced by the current EBM working group. In 2000 Sackett declared that he would never again “lecture, write or referee anything to do with evidence-based clinical practice.”\textsuperscript{109} At the time he claimed that this was because he had acquired expert status and was beginning to become an authority figure – something he recognized to be at odds with the fundamentally anti-authoritarian promises of the earliest versions of EBM. Sackett left without critiquing EBM, but the difference in the content of his statements on qualitative research (in 1997) and those quoted earlier in this section (from 2005) indicate a divide between the more critically-oriented and inclusive version of EBM Sackett seemed inclined to promote, and what Upshur has referred to as the “profound move within EBM away from fostering critical thinking, and creating a dependency on pre-interpreted, prepackaged sources of health evidence.”\textsuperscript{110} This shift seems to have happened more recently – i.e. since Sackett left the scene – and has corresponded with an increased reliance on a strict interpretation of the evidence hierarchy, and very little attention to qualitative research. As noted earlier, there

\begin{flushleft}
107 Sackett and Wennberg, 1636.
108 Ibid.
\end{flushleft}
are only cursory, hand-wavy statements on qualitative research in most guides today. The tension between the original commitments of EBM: to critical evaluation and to the provision of evidence that is immediately useful in practice, has been settled (for the worse, I argue) in favour of evidence that is predigested and thought to be ready for application in practice. Second, Sackett’s position is not representative of the core EBM position because if it were it would indicate the need for changes to the evidence hierarchy and this has not occurred as of the most recent edition of the official EBM handbook in 2005. If, as Sackett suggests, we should make use of a wide diversity of research methods, since they are each best suited to different research questions, a hierarchical ranking of research methods would make no sense. I will discuss problem further in the last section of the chapter. For now it will suffice to say that Sackett’s views are unfortunately not representative of the EBM position on qualitative research today.

4.4. Conclusions on Quantifiability

A well-designed and carefully executed qualitative study attains only the lowest level of recognition, if any, within the current evidence hierarchy. This placement reflects some of the underlying assumptions of EBM identified in the previous chapter, including the desire to tie medicine more closely to a traditional conception of good science. These sorts of assumptions have already been called into question. Quantitative evidence provides support for the technical side of medical practice. But why would we think that evidence is only available to support this one element of medical decision-making? Medical decision-making, as we saw in the last chapter, is complex and involves a great number of variables. Disciplines as varied as the social sciences, nursing, peace and conflict studies, human resource management, and education have a lot to offer physicians in the way of research on communication techniques and methods for conveying empathy and care. Qualitative research has long been gaining ground as a legitimate and important form of inquiry into complex problems, and its research methods have advanced considerably over the last half-century. The decision to categorize qualitative research as lowest quality is a throwback to a scientific tradition that devalues qualitative studies as unscientific. This decision is unjustified. Further, insofar as the evidence hierarchy implies the irrelevance of qualitative research it misrepresents the nature of research.

---

111 Morse, Swanson, and Kuzel, The Nature of Qualitative Evidence.
5. Managing Uncertainty

5.1. Uncertainty in Medicine

Immanuel Kant identifies the ways in which our perception of uncertainty is magnified in situations where life and death hang in the balance: “If, in a given case, we represent ourselves as staking the happiness of our whole life, the triumphant tone of our judgment is greatly abated; we become extremely diffident, and discover for the first time that our belief does not reach so far.”¹¹² Physicians have to deal with the weight of this uncertainty every day in situations where critical life and death decisions are made. Diagnoses are uncertain: What if a symptom was missed? What is the test was a false negative? Treatments are uncertain: What if this treatment causes an adverse event in this patient? What if there is a better treatment I do not know about? And so on. It is undoubtedly attractive, when in a difficult situation, to think that certainty is possible: to think that you can figure out the right answer. After all, the alternative is making decisions about the lives of patients while being relatively uncertain about the outcome. The sort of paralysis that can affect a physician in this situation may be dangerous to patients in the long run. Enter EBM, which “deals directly with the uncertainties of clinical medicine.”¹¹³

5.2. Uncertainty and EBM

According to the proponents of EBM, reasoning from pathophysiologic principles is necessary but not sufficient for medical decision-making. This is because “the rationales for diagnosis and treatment, which follow from basic pathophysiologic principles, may in fact be incorrect, leading to inaccurate predictions about the performance of diagnostic tests and the efficacy of treatments.”¹¹⁴ Implicit in this attack on pathophysiologic reasoning is a claim that clinical research evidence (particularly that obtained from RCTs) is unlikely to be incorrect. While EBM proponents do not come out and say this, since it has an air of infallibility generally thought to be

¹¹² Kant, 825A/853B. Thanks to Ross Upshur for bringing this to my attention.
¹¹³ Evidence Based Medicine Working Group, 2424.
¹¹⁴ Guyatt et al., “Evidence-Based Medicine,” 2421. Emphasis added. See Chapter Four for a full examination of the claims with respect to pathophysiologic reasoning.
inappropriate in science, it is really this idea that research evidence gets us closer to certainty, or infallibility, which drives a lot of the literature. One of the founders of EBM, Murray Enkin, has said that he was motivated to participate in developing EBM in part because it seemed to offer solid ground for decision-making in medicine. It seemed as though EBM would finally move physicians past uncertainty and beyond bickering over best practices. Just as Fisher originally promised, randomized experiments would end the need for debate and eliminate the need for interpretation and judgement. Members of the EBM working group counted on the rhetorical power of scientific evidence to stop disputes and settle longstanding questions. They were surprised to learn that evidence was not enough to end debates in medicine.

### 5.3. Conclusions on Uncertainty

C. David Naylor writes of the “Malthusian” growth of uncertainty that occurs when multiple technologies are combined in practice:

Take two technologies and they can be used in two different sequences; take five, and the number of possible sequences is one hundred and twenty. Furthermore, the elements in a clinical strategy are usually tested in separate studies, leaving few data on the chains of conditional probabilities that link sequences of tests, treatments, and outcomes. The play of uncertainty can be shown quantitatively when formal decision analysis is used as a tool to compare clinical strategies. The outcome is often either a ‘toss-up’ that rests squarely in the grey zones, or highly dependent on assumptions about one or more poorly defined variables in the model.

In the medical context, with the variety of human responses to treatment and the variety of combinations of treatments, the uncertainty grows exponentially. And, contrary to expectations, sociologists Stefan Timmermans and Marc Berg argue that uncertainty has not decreased as a result of EBM,

Based on interviews with residents in evidence-based medicine programs, we show that the incorporation of EBM did not remove clinical uncertainty and the reliance on experience in decision making but instead reshuffled knowledge hierarchies and introduced new kinds of uncertainties.

---

115 Personal communication with Murray Enkin.
117 Naylor, 840.
While proponents of EBM might lament this outcome, I think it was inevitable. While some increase in certainty might be possible, and should be pursued when possible, it was rather ambitious to suggest that the evidence hierarchy would put an end to clinical uncertainty. Moreover, there is some danger that the remaining uncertainty in decision-making could be ignored or downplayed by physicians who prefer to project an aura of confidence and certainty. It is a relief that this sort of obfuscation has not been widespread, since it would not have been justified, and would certainly have contributed in the long run to a crisis of trust in medicine. Valerie Miké has proposed an “ethics of evidence” for health professionals. This proposal consists of two injunctions: first, medical professionals are advised to create, disseminate and make use of scientific evidence, and second, medical professionals are impelled to recognize the irreducibly uncertain nature of evidence. Her emphasis is on the latter of the two principles, since it is more frequently neglected. I support these principles and argue that greater attention to the uncertainty in medical decision-making is important for the advancement of medicine.

The EBM hierarchy ranks most highly those methods thought to produce the most definitive and unquestionable results. This is meant to increase the certainty of physicians in the outcomes of particular treatments. Unfortunately, the assumption that methods higher in the hierarchy produce more certain evidence and lead to more certain medical decisions has not been borne out in practice.

6. **Is a Hierarchy Necessary?**

It is will always be useful to have clear, explicit criteria that demarcate a good RCT from a bad RCT, a good-quality cohort study from a poor-quality cohort study, and a well-conducted qualitative study from a poorly-conducted qualitative study. This is not what is at issue when we ask whether the evidence hierarchy of EBM is justified. What is at issue is whether it is possible to categorically rank an RCT against, for instance, a cohort study, or a case-control study against a qualitative study. I have argued against each of the five justifications offered for the hierarchy advanced by EBM. I turn now to the question of whether any hierarchy of evidence could be justified in medicine.
6.1. Different Questions, Different Methodologies

“I checked it very thoroughly,” said the computer, “and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you’ve never actually known what the question is.”\(^{119}\)

The proponents of EBM claim that they are not trying to say that the best way to answer every question is to conduct an RCT, just that RCTs should be performed whenever possible. Yet the evidence hierarchy seems to imply a categorical judgement about study quality regardless of the question under investigation. I argue that RCTs are not always better than other designs, even when RCTs are possible. Some important research questions cannot be answered by RCTs, and it would be inappropriate to allow the hierarchy to dictate the questions investigated in medical research even if there are clear reasons for choosing a particular study design to match a particular type of question.

Observational trials often investigate social, environmental, and lifestyle changes that could not otherwise be investigated in an RCT. For instance, if we want to know whether smoking causes cancer we are unable to randomize people into two groups – smoking and non-smoking – since people will not willingly take up or give up smoking according to the dictates of the trial design. Important lifestyle choices are not amenable to randomization. Further, it would not be possible to blind study participants to the intervention since it will be obvious who is and is not smoking. Thus the best possible study design for answering the question, “Does smoking lead to lung cancer?” is a prospective cohort study. Similarly, qualitative research is valuable because, when it is done properly, it answers questions that could not be answered by quantitative methods. If we are interested in asking why questions, such as, for instance, “Why do people decide to start smoking?” we require research methods such as surveys, interviews, and focus groups. If the question we want answered is this one, it makes no sense to say that an RCT is the best trial design. The best research method is that which provides the best answer to the question being asked.

Any hierarchy that attempts to rank research methods against each other will end up prioritizing certain research questions over others, thus restricting the scope of investigation.

\(^{119}\) Adams, 156.
This restriction is detrimental to medical knowledge and, by extension, medical practice.\textsuperscript{120} Further, any hierarchy presumes to have established the most important questions to be asked in medicine. The current evidence hierarchy, for example, is built on assumptions about best evidence, where such evidence is characterized by its precision, random allocation of subjects, quantifiability, lack of bias, and certainty. Yet even if the evidence produced by RCTs were to have these characteristics, it is not clear why we are choosing to determine what matters in medicine according to what sort of knowledge we are currently best at producing. Medicine should be driven by its cognitive goals and practical ends, not by what is easiest to investigate. The research questions valued most in medicine should be those that ask the most pressing questions, not those for which we have the most rigorous methods. Otherwise we are driven by what we are capable of producing rather than what we need to know. To use an analogy, consider a farmer who produces wheat because she has the best equipment for harvesting wheat (the latest combine, perhaps) when her family and the local community are in need of potatoes, corn and peas. It is tempting to do what we (ostensibly) have the best tools to do, but we are better served by staying focused on what it is we actually need. What it is that medicine needs is a matter for deliberation. The goals of medicine should not be dictated by our research methods.\textsuperscript{121}

6.2. Methodological Pluralism

In light of a lack of justifications for the current evidence hierarchy, and for evidence hierarchies generally, it is worth discussing possible alternative ways of organizing evidence. Henrik Wulff, for instance, advocates viewing EBM “not [as] a new discipline but [as] one ingredient in the production of good clinical care.”\textsuperscript{122} Petticrew and Roberts have suggested the importance of “horses for courses” with respect to research methods, and proposed a “matrix” of evidence.\textsuperscript{123}

\textsuperscript{120} I discuss these issues further in Chapter Six.\textsuperscript{121} Hanson and Callahan have recently edited a volume on The Goals of Medicine, which is the product of a four-year project initiated by the Hastings Center involving working groups of health professionals, philosophers, social scientists, politicians, theologians and legal professionals from fourteen countries. While a diversity of perspectives are offered in the volume, there is a consensus on certain core goals of medicine: 1) the prevention of disease and injury and the promotion and maintenance of health, 2) the relief of pain and suffering caused by maladies, 3) the care and cure of those with a malady, and the care of those who cannot be cured, and 4) the avoidance of premature death and the pursuit of a peaceful death.\textsuperscript{122} Wulff in Daly, 36.\textsuperscript{123} Petticrew and Roberts, “Evidence, hierarchies and typologies.”
Muir Gray has developed a “typology” in order to schematically indicate the different strengths of each of the research methods.\textsuperscript{124} Bluhm has proposed a “network” of evidence that pays closer attention to the interdependence of evidence at each level of the traditional hierarchy.\textsuperscript{125} There are no shortages of different names for what amounts to a renewed appreciation for methodological pluralism in medical research. Recognition of pluralism forces physicians to become more aware of the strengths and weaknesses of different research methods, and though more challenging in practice, at least discourages the false complacency that comes with rigid hierarchies.

Outspoken EBM critic Ross Upshur has provided a model of evidence that aims to capture the variety of meaningful types of evidence in medicine as well as their interrelationships, without imposing any comparative ranking. The four distinct but related concepts of evidence are: “qualitative/personal”, in which evidence is narrative in nature, socially and historically context-specific and individualized; “qualitative/general” in which evidence is social, historical and general; “quantitative/general” in which evidence is statistical, general, impersonal and quantitative; and “quantitative/personal” in which evidence is quantitative, yet individualized. Of these, the emphasis of EBM has been on quantitative/general evidence, often at the expense of other types of evidence. Exclusive focus on this one type of evidence is unjustified. This model acknowledges the value of each type of evidence and provides a general framework for understanding how each type of evidence can complement the others. This does away with hierarchies altogether and embraces a more complicated – yet more accurate – description of the plurality of useful research methods.\textsuperscript{126}

Nancy Cartwright has recently, and provocatively, described the function of evidence-based standards as follows:

Grading schemes [such as the evidence hierarchy] don’t combine evidence at all – they go with what’s on top. But it seems to me to be daft to throw away evidence. Why are we urged to do it? Because we don’t have a good theory of exactly why and how different types of evidence are evidence and we don’t have a good account of how to make an assessment on the basis of a total body of evidence. Since we don’t have a prescription for how to do it properly, we are urged not to do it at all. That seems daft too. But I think it is the chief reason that operates. That is why the philosophical task is so important.\textsuperscript{127}

\textsuperscript{124} Muir Gray, \textit{Evidence-based Healthcare}.
\textsuperscript{125} Bluhm, “The ‘hierarchy of evidence’ and the structure of medical research.”
\textsuperscript{126} Upshur et al. “Meaning and Measurement.”
\textsuperscript{127} Cartwright et al., “Evidence-based Policy,” 15.
The philosophical task Cartwright refers to is that of developing a theory of evidence that is both philosophical and practicable. Traditional philosophical theories of evidence have tended to focus on how to determine whether something is evidence (usually in terms of the probabilistic relationship between evidence and hypothesis or on a further account of explanation). These accounts are impractical. Practical accounts of evidence implicit in current standards of evidence proposed by, for instance, the EBM movement, have focused on ranking different types of evidence as better or worse than others, but have neglected the more basic question: what makes something evidence in the first place? They are thus not comprehensive (or, in Cartwright’s terms, ‘philosophical’). This no doubt leads to confusion over the place of expert judgement, patient values and anecdotes as evidence – as was discussed in the last chapter. A theory of evidence should guide us in determining whether anecdotes and expert judgements are evidence at all rather than assuming they are not because they fail to meet arbitrary standards of good evidence captured by the assumptions underlying the evidence hierarchy.

Cartwright’s proposal is ambitious. In line with the contextual emphasis of CCE, I believe such a theory is impossible on the grand scale. Cartwright seems set on having a grand theory of evidence, but this simply repeats the mistakes made by proponents of EBM, who tried to offer a general theory of evidence and ended up losing touch with the contextual demands of medicine. I believe that a general theory of evidence is impossible and that attempts to provide such a theory will be ultimately detrimental to our ability to contextualize the nature of evidence in particular settings. A context-specific account of evidence, however, would be a significant addition to the medical debate. While I do not develop such an account in this dissertation, I do outline several of the social institutions that need to be in place for discussion on such an account to proceed (see Chapter Seven for further details). To return to the subject matter of this chapter, Cartwright’s discussion reiterates the “daftness” of proposals to rank research methods against each other and the need for a less restrictive and non-hierarchical description of medical evidence.

7. Summary

---

128 See for instance Achinstein.
129 In Cartwright’s terms, what makes something candidate evidence?
The CCE framework adopted in Chapter Two stresses the importance of critically scrutinizing background assumptions. According to my analysis in Chapter Three, there are three foundational background assumptions underlying EBM. In this chapter I examined the third of these: the assumption that a hierarchy of evidence is necessary to guide medical research and practice. This assumption was supported by five more detailed, substantive assumptions, which I identified at the outset of this chapter. The evidence hierarchy ranks research methods according to their ability to produce evidence that is: precise, randomized, unbiased, quantified and relatively certain. The particular version of the evidence hierarchy developed on the basis of these assumptions about best evidence is only justified insofar as these characteristics really do demarcate good from bad evidence. In this chapter I have argued that each of these justifications fails. As a result of these collective failures, and in the absence of further justifications from EBM proponents, the attempt to justify this version of the evidence hierarchy fails. Moreover, any such attempt is incoherent. Different research questions require different research methods. Any evidence hierarchy is necessarily an attempt to rank the most important questions in medicine, and this – even if possible – should not depend on characterizations of good evidence but on the goals and ends of medicine. Otherwise medicine will be driven by what is possible within current methods, rather than what is required by practice or most important according to its philosophy of health and disease.
Chapter Six

The Future of Evidence-based Medicine: Diversity and Evidence-based Complementary and Alternative Medicine

“[O]nly through diversity of opinion is there, in the existing state of human intellect, a chance of fair play to all sides of the truth.”

According to critical contextual empiricism (CCE), diversity contributes to the production of knowledge by upholding and advancing the objectivity of inquiry. In light of this commitment – made more explicit in my version of the CCE norms – we have reason to ensure that the medical community adopts rules and practices that encourage diversity and responsiveness to diversity. In this chapter I investigate whether the evidence-based medicine movement moves us toward or diverts us from these sorts of rules and practices. I argue that the EBM movement suppresses diversity in medicine. This is manifested in two ways. First, the centerpiece of the EBM movement – the hierarchy of evidence – overstates the epistemic value of the randomized controlled trial. This diverts the attention of both researchers and physicians from valuable evidence at lower levels of the evidence hierarchy such as that produced by cohort, case-control and qualitative studies, and decreases the variety of both medical research questions and corresponding treatments. Thus, EBM suppresses methodological diversity through internal constraints on the mainstream medical community. Second, increased social and political pressure resulting from EBM constrains diverse perspectives offered by those outside of the medical mainstream. Researchers in complementary and alternative medicine, for instance, have been instructed to conform to current standards and produce RCTs if they hope to gain credibility. This demand for conformity persists despite concerns raised about the limits of the RCT, the value of lower-ranked research methods, and methodological innovations proposed by researchers in complementary and alternative medicine. In other words, EBM also provides a

---

1 Earlier versions of sections 1-3 of this chapter were published as: Borgerson, “Diversity and Objectivity in Medicine”; earlier versions of sections 4-9 of this chapter were published as: Borgerson, “Evidence-based Alternative Medicine?”
2 Mill.
3 I explain and defend this argument in Chapter Two.
4 Meta-analyses of RCTs are meant to be included by this statement. A full defence is offered in Chapter Five.
number of *external* constraints on methodological diversity. Moreover, these external constraints also diminish the presence of diverse metaphysical accounts of health and disease such as those underpinning various CAM approaches. Thus EBM suppresses the development of alternative theoretical frameworks or what I will call theoretical diversity. In the CAM context, the methodological and theoretical issues are often intertwined.

In what follows, I begin with a brief discussion of the internal constraints on methodological diversity, which build on arguments in the previous chapter. I then turn to a more comprehensive analysis of the ways in which EBM inhibits the development of both methodological and theoretical diversity through external constraints. To this end I examine current debates in complementary and alternative medicine (CAM). In order to regain a healthy attitude toward diversity in medicine, I argue that mainstream medical researchers should pay attention to the developments in complementary and alternative medical research. I outline a number of CAM innovations and discuss some of the ways in which CAM research provides a vital critical perspective on the background assumptions of EBM.

1. **Internal Constraints on Methodological Diversity**

The EBM movement is, at its core, an attempt to make medicine more standardized. It is no surprise, then, that the EBM approach contributes to a restriction on diversity within the domain of medicine. As proponents of the EBM movement have freely admitted, the movement has led to a somewhat zealous support for RCTs and a near-dismissal of all other sources of evidence. Proponents of EBM “almost succeeded in colonizing the meaning of the word evidence, so that it becomes synonymous with that evidence produced by a randomized controlled trial.” Because researchers face various community incentives to publish research, and require funding from agencies that uphold the EBM hierarchy, researchers have been compelled to produce RCTs whenever possible. And because medical guidelines are based largely (if not fully) on the results of meta-analyses and systematic reviews of RCTs, there is little incentive for researchers to bother producing research based on any of the other methods. Of course, these other research

---

6 These are *external* constraints because they are placed on members of other communities.
7 This could also be called metaphysical or framework diversity.
8 Timmermans and Berg.
9 Daly.
10 Ibid., 231-2.
methods have not been totally abandoned, but EBM has certainly diverted researchers from using a variety of study designs.  

In addition to the arguments I have already made about the shortcomings of the evidence hierarchy, I would like to highlight the impact of any evidence hierarchy on methodological diversity. Because every research method is able to address a specific range of questions, the diversity of research questions decreases whenever research methodologies are limited. This bears repeating. The range of research questions that can be asked and legitimately pursued within the mainstream medical community is circumscribed by the EBM hierarchy of evidence. This, after all, is essentially the point of any hierarchy of evidence: focusing attention on certain types of evidence at the expense of others. The result, in this case, has been a significant decrease in the production, publication and institutional influence of studies ranked in the middle and near the bottom of the hierarchy. A narrow range of research questions corresponding to one or two best research methodologies means a narrow range of research results. This has immediate implications for practice. The evidence hierarchy decreases attention to certain methodological designs and thus limits the range of possible treatments available to patients. Failure to appropriately value the contributions made by a variety of research methods leads to a decrease in methodological diversity within mainstream medicine.

2. External Constraints on Methodological and Theoretical Diversity

There is a great deal of social, economic and political power behind the EBM movement. This power not only influences the research patterns within mainstream medical research, as noted above, but also places tremendous pressure on researchers outside the medical mainstream to produce what it deems top-ranked evidence. Most complementary and alternative medical researchers – like those in mainstream medicine – wish to evaluate their therapies using methods

---

11 The term “randophilia” was coined by Alvan Feinstein to capture the obsession with randomized trials that took place with the introduction of EBM. Feinstein was one of most outspoken critics of EBM in the 1990s. Ibid., 104.  
12 To be more specific, the current evidence hierarchy valorizes methodologies that are best suited to study interventions such as pharmaceutical drugs (which are easiest to placebo-control), and least well suited to study behavioral, environmental, and other lifestyle or preventative interventions. This problem is discussed later in this chapter.  
13 “[W]e get little new understanding of the barriers to change in relation to either patients or practitioners, and the interventions designed for overcoming these barriers continue to be hit or miss. This is a case where we need diversification of research methods, and other evidence.” Ibid., 232.
that are rigorous and that are consistent with their philosophies of medicine and healing. These investigators have three ways to relate their work to EBM. They can accept the EBM hierarchy and carry out RCTs when possible; they can accept the EBM standards but argue that the special characteristics of alternative medicine warrant the acceptance of ‘lower’ forms of evidence; or they can challenge the EBM approach and work to develop new research designs and new standards of evidence that reflect their approach to medical care. For several reasons, this last option is preferable. First, it will best meet the needs of alternative medical practitioners. Moreover, because similar problems beset the evaluation of mainstream medical therapies, reevaluation of standards of evidence will benefit everyone in the medical community – including, most importantly, patients.

3. What is Complementary and Alternative Medicine?

What is considered to be mainstream medicine in North America today is the result of social, political and economic forces in the past century. The boundaries of mainstream medicine do not correspond with any formal distinction between science and pseudo-science. For this reason, and because mainstream medicine is continually evolving to include and exclude different practices and therapies, it is difficult to accurately define complementary and alternative medicine. Because CAM is a catch-all term used to refer (loosely) to those practices outside the mainstream, and is therefore highly context dependent, the National Institutes of Health (NIH) defines complementary and alternative medicine as:

> All health systems, modalities and practices and their accompanying theories and beliefs, other than those intrinsic to the politically dominant health system of a particular society or culture in a given historical period.\(^4\)

This allows traditional Chinese medicine (TCM) to be a mainstream approach in China, where it is practiced widely in mainstream hospitals. TCM is, on the other hand, alternative in Canada. This broad definition, as applied in North America, encompasses biologically-based treatments (herbs, special diets, and vitamins), manipulative and body-based treatments (chiropractic, massage, and osteopathy), energy therapies (reiki, magnet therapy, and qi gong), mind-body

\(^{14}\) National Institute of Health Panel on Definition and Description. “Defining and describing complementary and alternative medicine.”
Alternative medicine has a tremendous amount of public support in North America. Americans spent between $37 and $47 billion on alternative medicine in 1997, and these numbers are growing. According to the most recent survey released by the U.S. National Center for Health Statistics (NCHS), approximately 36% of American adults are currently using some form of alternative medicine. Similar polls in Canada indicated that between 42 and 50% of the population have used some form of alternative medicine in the past year. This was a more than 80% increase when compared with a poll conducted five years earlier. Despite what critics continue to regard as a serious lack of scientific evidence, alternative medicine appears to be gaining acceptance throughout North America.

4. Expectations from Mainstream Medicine

In an influential clinical epidemiology textbook, Sackett and colleagues argue for the increased use of research evidence in practice. One of the reasons they offer is that “to base your treatments of commonly encountered problems on the advice of some ‘expert’ who publishes treatment recommendations but no supporting evidence puts you on par with the barefoot doctor.” According to Sackett and colleagues, one can avoid sinking to the level of the barefoot doctor by shifting one’s alliance to the “scientifically rigorous” methods of the randomized controlled trial. As this position makes apparent, the tension between mainstream and alternative medicine plays no small part in the debate over EBM. Two articles in top medical journals typify the mainstream reaction to alternative medicine:

There is no alternative medicine. There is only scientifically proven, evidence-based medicine supported by solid data or unproven medicine, for which scientific evidence is lacking. Whether a therapeutic practice is “Eastern” or “Western,” is unconventional or

---

15 These terms are used interchangeably.
17 Angus Reid, “Canadian poll details ‘Alternative Medicine’ use.”
18 Trends are similar in Europe.
20 Ibid., 194.
mainstream or involves mind-body techniques or molecular genetics is largely irrelevant except for historical purposes and cultural interest.²¹

It is time for the scientific community to stop giving alternative medicine a free ride. There cannot be two kinds of medicine—conventional and alternative. There is only medicine that has been adequately tested and medicine that has not, medicine that works and medicine that may or may not work. Once a treatment has been tested rigorously, it no longer matters whether it was considered alternative at the outset. If it is found to be reasonably safe and effective, it will be accepted. But assertions, speculation, and testimonials do not substitute for evidence. Alternative treatments should be subjected to scientific testing no less rigorous than that required for conventional treatments.²²

According to these powerful demands, alternative medical researchers should strive to conduct the same sorts of clinical trials currently regarded as the gold standard of medical research. Journal articles such as “A warning to complementary medicine practitioners: get empirical or else,” in which CAM researchers are advised to start producing RCTs, convey these expectations in an admirably candid way.²³ CAM research must, at the very least, meet current standards; in some cases, CAM research is even required to exceed current standards, based on certain Bayesian ideas that advise extra scrutiny for those practices and treatments that have low prior probabilities.²⁴ Conformity to mainstream research methodologies is not requested, but demanded. And these demands are being heard.

5. The Sameness Approach²⁵

The push for evidence-based CAM is reflected in the CAM literature and in the recent creation of journals such as Evidence-based Complementary and Alternative Medicine, which are seen as

---

²¹ Fontanarosa and Lundberg, “Alternative medicine meets science,” 1619.
²² Angell and Kassirer, 839.
²³ Haynes, “Commentary: a warning to complementary medicine practitioners: get empirical or else,” 1632.
²⁴ Of course, alternative medical research is not alone in failing to live up to current standards of evidence. As Morreim points out, “Much of actual clinical practice does not and never can measure up to the scientific standards to which critics of CAM would like to hold alternative modalities” (222). A large percentage (roughly estimated around 80%, though the exact figures are highly contested) of medical treatments currently offered in hospitals and clinics across North America have never been tested by an RCT. Research into alternative medicine is required to meet the highest standards even though many currently accepted medical practices have not met (and may never meet) those same standards. University of Sheffield.
²⁵ Characterizing the choices as the “sameness” and “difference” approaches reflects the influence of Catharine MacKinnon on this chapter. MacKinnon, Toward a Feminist Theory of the State. An earlier version of this chapter was published as a working paper and this earlier version retains more of the explicitly feminist framework: Borgerson, “Playing by the Rules.”
an important step forward for CAM research. The research department at the Canadian College of Naturopathic Medicine performs RCTs, meta-analyses and systematic reviews almost exclusively. In an article summing up CAM perspectives on research, Fitter and Thomas reinforce these expectations of mainstream medicine, “The pervading view that a classical RCT will always be the first-choice methodology for any evaluation means that if, for any reason, it cannot be used, the resulting study is judged to provide less persuasive findings: they will be second best.”

There are a number of advantages to playing by the rules of EBM. EBM has been widely accepted in medical schools and research institutes around the world as the best approach to medical decision-making and as an excellent guide for the assessment of research methodologies. As such, it influences the evaluation of studies by medical journals and the allocation of research dollars by funding agencies. Alternative medical researchers could very easily concede the legitimacy of this approach and try to work within it. If researchers in alternative medicine are motivated by a desire to earn legitimacy for what they believe to be excellent and often overlooked treatments, they may be motivated to become “evidence-based”. Alternative medical researchers may also concede to mainstream demands out of a desire to be included under medical insurance plans or, in the case of members of the nutraceutical industry, by a desire to influence the prescribing power of physicians and practitioners.

Some alternative medical researchers have taken the demands for “gold standard” research seriously and, especially in the last decade or so, there have been many studies evaluating a variety of alternative medical treatments. Some results have been negative; for example, in two studies published in the 1998 special issue of JAMA dedicated to alternative medicine acupuncture was found to be no more effective than placebo in relieving pain caused by HIV-related peripheral neuropathy and Garcinia cambogia failed to produce significant weight loss. On the other hand, the use of moxibustion (burning herbs at an acupuncture point)

---

27 Personal correspondence with Kieran Cooley, ND, Assistant Professor and Research Trial Coordinator at the Canadian College of Naturopathic Medicine (CCNM).
28 Fitter and Thomas, “Evaluating complementary therapies for use in the National Health Service,” 91.
29 Nutraceuticals are food substances that are thought to have health benefits. For example, red wine and garlic are believed to have antioxidant effects. Many herbs are also included in this category since they are often thought of as both foods and drugs.
30 Shlay et al. “Acupuncture and amitriptyline for pain due to HIV-related peripheral neuropathy: a randomized controlled trial.”
to correct breech presentation in late pregnancy, TCM herbs for irritable bowel syndrome, glucosamine for the treatment of osteoarthritis, and acupuncture for nausea and vomiting, among others, have been shown to be effective in RCTs. In addition, the Cochrane collaboration has, as of July 2007, produced meta-analyses of 275 CAM treatments, which pooled over 7,000 individual RCTs. While many of the meta-analyses indicated no effect of the treatment under investigation, a good number were positive. There are plenty of examples of alternative medical research that has adhered to the standards outlined by EBM. Many of these successful treatments, despite achieving the status of best evidence, have not been accepted into mainstream medical practice, although there is some indication that medical students and new physicians are taking more of an interest in certain alternative treatments. The vast majority of alternative medical treatments, however, are not even at this preliminary stage of acceptance.

6. The Difference Approach

While many CAM communities have begun to adopt evidence-based rules and practices, members of a vocal minority resist this assimilation. These practitioners and researchers argue that because they practice in a “different paradigm”, from a different world view, and hold a different philosophy, they should not be held to the same standards as “Western” medical research. In essence, they accept the EBM hierarchy as legitimate for scientific evaluation but argue that special exceptions should be made for alternative medical research because it is different and cannot meet the standards of EBM. Even with increased emphasis on RCTs today, there are CAM researcher who prefer to use lower-ranked research methods. According to


31 Heymsfield et al. “Garcinia Cambogia (hydroxycitric acid) as a potential antiobesity agent: A randomized controlled trial.”
32 Cardini and Weixin, “Moxibustion for correction of breech presentation: A randomized controlled trial.”
34 McAlindon et al. “Glucosamine and chondroitin for treatment of osteoarthritis: A systematic quality assessment and meta-analysis.”
35 National Institutes of Health. “Acupuncture.”
36 The Cochrane Collaboration. A survey of these meta-analyses, in 2004, indicated that, of the 145 reviews available at the time, “the largest number of reviews were classified as insufficient evidence (n=82, 56.6%); followed by positive effect (n=36, 24.8%); possibly positive effect (n=18, 12.4%); and no effect (n=7, 4.8%). One each was classified in the other two categories [equal effect of two treatments, and harmful effect].” Manheimer et al., “Cochrane reviews of complementary and alternative therapies: evaluating the strength of the evidence.”
37 Two examples of positively reviewed CAM treatments are: Herxheimer and Petrie, “Melatonin for preventing and treating jet lag.” and Pittler and Ernst, “Horse chestnut seed extract for chronic venous insufficiency.”
38 Maha and Shaw.
proponents, alternative medical treatments are often highly individualized (as compared to the
generalized treatments offered in an RCT), complex (have a number of therapeutic components),
require physical treatment to which it is difficult to “blind” patients and practitioners (for
example, needling in acupuncture), actively involve the therapist and patient as integral parts of
treatment (making randomization and standardization difficult), set different (and multiple) end
points based on different philosophies of medicine, and rely on principles of self-healing and
mind-body control (which incorporate, rather than rule out, the placebo effect). These and other
problems, including social and economic difficulties in funding and organizing large-scale
studies on often un-patentable treatments, are often raised as explanations for the persistence of
lower ranked research methods in alternative medicine.

In response to these appeals to difference, critics decry the “quackery” of alternative
medicine and caution patients to avoid alternative medical practitioners. Even people who are
generally sympathetic to alternative medicine respond negatively to this sort of approach. A
representative response is offered by Wayne Jonas, “Claiming that their practices are too
‘individual’ or ‘holistic’ to study scientifically, many alternative medical practices hide behind
anecdote, case series, or ‘outcomes’ research.” Notice here that anything short of full
commitment to the EBM hierarchy is provoking this response – even observational research is
not enough. (This illustrates the elevation of the RCT within the medical community and the
clear demands for RCT evidence placed on alternative medical researchers.) Alternative medical
researchers have found that there is no guarantee that labeling yourself different will earn you
any respect from the mainstream medical community. In fact, it is just as likely, if not more
likely, to lead to worse treatment. In the eyes of the mainstream medical community, there is an
epistemological standard that alternative medicine simply fails to meet. When it is assumed that
everyone should meet the same standard, an appeal to difference comes across as a sign of
weakness – an indication of failure and an appeal for special treatment.

7. What Gets Lost in Either Approach

39 Anthony, “Some methodological problems in the assessment of complementary therapy,”
40 Jonas, “Alternative medicine – Learning from the past, examining the present, advancing to the future,” 1616-17.
41 MacKinnon, Toward a Feminist Theory of the State.
Both of the approaches just outlined have shortcomings. In what follows, I briefly discuss the shortcomings of the difference approach before turning to an in-depth critical analysis of the sameness approach.

### 7.1. The Difference Approach: Empirical Evaluation

The difference approach is more obviously problematic than the sameness approach. When CAM researchers rest on their differences and request special treatment, this undermines many of the motivations that alternative health researchers begin with – including the desire to gain legitimacy for alternative treatments and the desire to become part of the scientific community in order to discover which alternative treatments “really work.” There are good reasons for anyone to accept at least the most basic principles of scientific research, including empirical testing and attempts at falsification of hypotheses. Even though they may not translate directly into the evidence hierarchy, there are some basic elements of good evidence (trial-and-error empiricism, for example) that can be agreed upon and should be upheld. As long as there are con artists with an interest in selling bogus remedies there will be a need for ways of distinguishing good treatments from harmful ones. Anyone truly dedicated to the health of others will recognize the value of having mechanisms for determining which treatments are effective and which are not. Applying for special exemptions from the demands of EBM could be tempting, but requires that researchers write off the value of the scientific endeavor – a decision that, for reasons far beyond those just outlined, is far from ideal.

### 7.2. The Sameness Approach: Methodological and Theoretical Diversity

The social influence of EBM contributes to a decrease in the diversity of competing theories by requiring the assimilation of alternative theories of health, and alternative methodologies, to the mainstream theoretical framework. In order to demonstrate the loss of theoretical and methodological diversity that can occur when CAM researchers blindly adhere to EBM standards I will outline the challenges faced by CAM researchers as they attempt to design studies that are both scientifically rigorous and relevant to their medical practice. To narrow the focus, I consider researchers in naturopathic medicine. Although homeopaths and TCM practitioners may face
distinct problems that are not fully incorporated into this discussion, many of the challenges faced by naturopaths (the “GPs” of alternative medicine) are likely to be similar to those encountered by other alternative systems.

7.2.1. Case: Naturopathic Medicine

The guiding principles of naturopathic medicine are:

1. Do no harm
2. Identify and treat the causes
3. View the doctor as a teacher
4. Treat the whole person
5. Emphasize prevention
6. Support the healing power of the body

Naturopathic doctors are directed to treat each patient by taking into account individual physical, mental, emotional, environmental, and social factors. The patient is a vital part of the healing process and is encouraged to take personal responsibility for his or her own health. Treatments are offered that enhance and support the healing power of the body. Naturopaths employ a selection of treatment modalities usually including Asian medicine/acupuncture, botanical (herbal) medicine, clinical nutrition, homeopathy, lifestyle counseling, and physical medicine (massage, ultrasound, hydrotherapy, etc.).

Researchers and practitioners in naturopathic medicine encounter a number of challenges when attempting to follow the EBM approach. Some of these difficulties are shared by mainstream physicians. For instance, because of the naturopathic commitment to highly individualized care, there are a number of challenges in applying the results of highly generalized evidence to patients. The problem, which may be somewhat more noticeable in naturopathic medicine because of the number of patients with chronic or multiple conditions, is that “excellent evidence does not necessarily translate into excellent or successful therapy.”

We saw concerns of this type raised by mainstream physicians in Chapter Four. Similarly, the naturopathic commitment to “view the doctor as a teacher” highlights the importance of listening.

---

42 Canadian College of Naturopathic Medicine (CCNM).
43 Upshur, “Seven Characteristics of Medical Evidence,” 94.
to and educating patients, rather than taking an overly paternalistic approach to their care. Because of the expectation that teachers are themselves lifelong learners, this commitment highlights the importance of shared decision-making in naturopathy. Concerns that the EBM approach is in tension with shared decision-making were discussed in Chapter Four. In addition to the extensive overlap in the concerns of naturopaths and mainstream physicians on these points, there are two further naturopathic principles that add new perspectives to the assessment of EBM.

7.2.1.1. “Treat the Whole Person”

The naturopathic principle, “treat the whole person,” reflects a commitment to providing treatments which address as many of the complex causes of illness and disease as possible. Because of this, naturopaths often recommend a variety of treatments for a single condition, from nutritional changes to physical treatments such as acupuncture. In tension with this multifaceted approach to care, naturopaths note the trend toward pharmaceutical treatments for all illnesses which is currently underway in mainstream medical research and practice. Mainstream physicians are, more than ever before, prescribing drugs as the treatment of choice for any presented condition. It would be surprising if this was unrelated to the type of research being emphasized in the medical community. Research questions that fit within the RCT design are limited to variations of: ‘is treatment x effective for condition y’ where x is a single variable and y is a specific outcome. The RCT is well suited to the evaluation of drugs. As Wendy Rogers has written,

[The EBM] way of defining problems lends itself to a research agenda in which the immediate and identifiable causes are investigated and treated. The focus is upon physical interventions acting on the diseased person, such as taking medicines to kill infectious agents.

This is supported by extensive anecdotal evidence, as well as studies such as that by Everitt, Avorn, and Baker, in which general practitioners were presented with the hypothetical case of an elderly patient suffering from insomnia (one of the most common symptoms in outpatient

---

44 Angell.
45 Rogers, “Evidence-based Medicine and Women,” 68.
medicine), and “despite many possible non-pharmacologic therapies for the patients presented, 46% of physicians identified a prescription medication as the single most effective therapy.”

The population of North America is one of the most heavily medicated in the world (some would suggest “overmedicated”). Making judgements about the care of individual patients on the basis of RCT research alone may contribute to this problem. Insofar as EBM encourages physicians to treat on the basis of evidence from RCTs, it is also implicated in this growing problem.

Consider the following situation. A mainstream physician practicing EBM encounters a patient who is suffering from asthma and who is in need of treatment. Upon formulating a question, completing a literature search on MEDLINE and narrowing the results to meta-analyses of RCTs, she discovers that corticosteroids are highly effective for the treatment (that is, symptom management) of asthma. We see here both the limitations and significant advantages of the evidence-based approach. The physician was able to find good quality evidence for the effectiveness of a treatment very quickly, and with minimal hassle. The evidence hierarchy allows physicians to begin to order and organize what might otherwise be an overwhelming amount of information. It also guides the physician to the best RCT evidence – so he or she would no longer consider prescribing outdated treatments for asthma, for instance.

Unfortunately, the reliance on this hierarchy also means that the physician overlooks the studies lower in the hierarchy that also attempt to answer the question, “What treatment is effective for asthma?” though in a slightly different way, or perhaps in different contexts or settings.

Given that RCTs are designed to mimic a closed, controlled setting in which only one variable is studied and neither the physician nor patient are aware of which treatment is being administered, it cannot be used to evaluate certain types of treatments. Because social and environmental systems are open and dynamic, different types of research methodologies have been designed for evaluating treatments in those settings. Observational research on treatments for asthma might have yielded results indicating that asthma rates are affected by environmental conditions such as intense heat and smog, and that moving to different environmental settings can be very helpful for the management of asthma symptoms. The research might also have suggested the value of lifestyle interventions such as changes in nutrition or exercise. This information was not examined by the physician in this example, nor did it affect the suggestions

---

46 Everitt, Avorn, and Baker, “Clinical decision-making in the evaluation and treatment of insomnia.”
47 Harvard medical professor and former co-editor of the New England Journal of Medicine, Marcia Angell, says, “In my view we have become an overmedicated society.” Angell, The Truth About the Drug Companies, 169.
made to the patient. If there is a meta-analysis of RCTs indicating that a pharmaceutical drug is effective in the treatment of asthma, and merely qualitative evidence suggesting the effectiveness of exercise or other lifestyle changes, the physician who adheres closely to the evidence hierarchy would be inclined to treat the patient according to the highest ranked evidence and prescribe the drug.

For asthma, the common sense treatment in a particular case may often be lifestyle changes such as quitting smoking and spending less time in a smoggy city during the summer. I am sure any conscientious physician would point out that the treatments beyond pharmaceutical drugs are common sense and that physicians are aware of them and do make a point of treating patients in a wide variety of ways. And I am sure this is true: overweight people are advised to adjust their diets and to increase their daily exercise, people with depression are counseled on life-style changes that may be helpful, and those with asthma are offered advice about how to adjust their environment. What it is critical to realize, though, is that insofar as physicians act in this way, they are relying heavily on the “judicious” and “conscientious” elements of the EBM definition. They are, in effect, bypassing the evidence hierarchy. In fact, this is precisely why the second definition of EBM had to be introduced. If physicians adhere too strictly to the EBM hierarchy, they might end up prescribing drugs all the time, even though life-style interventions are sometimes the better choice.\footnote{Orlistat (Xenical), for example, is the latest in pharmaceutical drugs designed to ‘treat’ obesity in spite of decades of observational evidence demonstrating the value of exercise and dietary changes. For extensive criticism of this and other lifestyle drugs see the consumer watchdog Public Citizen’s website at: www.worstpills.org.}

The introduction of the phrase “conscientious, explicit and judicious use of current best evidence” to the second EBM definition created a sort of loophole for clinicians. It expanded the scope of physician autonomy, in the hopes of responding to the concerns raised in medical debates about the cookbook nature of the original formulation. While these concessions were permitted, the hierarchy remained consistent throughout the various formulations of EBM. As a result, we are now in a position in which physicians who adhere strictly to the evidence hierarchy may end up over-prescribing pills as treatment for diseases rather than offering less invasive and often safer lifestyle interventions. On the other hand, if physicians rely on the “conscientious and judicious” interpretation of evidence, they will have a much wider range of options for treatment, more in line with the full array or network of evidence that is produced from a variety of sources and on a variety of research questions. But this looks peculiar. Rather than admitting the flaws
of the evidence hierarchy, advocates have kept it intact and offered loopholes to those physicians who feel the need to deviate from the dictates of best evidence. But this places far too much weight on the terms conscientious and judicious. These terms in the description of EBM save it from absurd conclusions (such as that the best treatment for obesity is generally a pharmaceutical drug) by allowing physicians to use common sense in the treatment of disease. But, from a theoretical standpoint, even with the addition of these loophole terms the evidence hierarchy is clearly not doing the task EBM proponents claim it is doing. It is not possible to identify the ideal method for all medical questions about treatment without inappropriately neglecting some treatments and treatment questions.

Making judgements about the care of individual patients on the basis of RCTs alone may contribute to overmedication simply because pharmaceutical drugs are most easily amenable to testing in RCTs. The evidence produced by RCTs is not always (or even usually) helpful at answering social, economic, or environmental questions related to health. Observational studies are better designed to answer these broader sorts of questions, and insofar as the EBM hierarchy downplays the value of these other sources of evidence, it seriously limits the scope of good research and the corresponding scope of recommended treatments. This is especially unhelpful within a holistic model of health. The social, economic, and environmental context of research is explicitly factored out within the EBM model. Because naturopaths have a commitment to looking closely at social, economic, and environmental factors when evaluating the health of patients, they are more aware of the dangers of the trend toward offering pharmaceutical drugs – or even standardized herbal remedies – for all medical conditions. However, there are reasons why all health professionals should be worried about this trend.

7.2.1.1.1. Context-stripping (RCTs): What Gets Missed

Ruth Hubbard first introduced the idea of “context stripping” to describe the process of detaching objects from the contexts in which they naturally occur in order to understand them. There is no doubt that this can be a very effective way of examining certain features of a given object. The problem with this approach is that it means we are unable to understand the ways in which the object normally functions in relation with other factors in its usual context. So we trade off

---

49 Hubbard, “Science, Facts, and Feminism.”
understanding one thing (how an object functions in relation to its usual context) for another (an understanding of the isolated object). Depending on whether we think there is any value in the first type of understanding, we might think twice about the assumption that context-stripping is always the best way to gain knowledge. In light of their holistic principles, naturopaths are among those who value the type of knowledge achieved by contextual investigations.

We see the dangers of context-stripping in areas of science most closely connected to technologies or practices. In medicine, for instance, we aim to understand the effects of treatments as if the patient were not there – in a sense what we try to do is erase the individual patient as much as possible so that we can get at data on the isolated treatment-effect pathway. There are many problems raised by this. First, because context does matter we end up with different results in different patients, though we try to average these out in final calculations. Second, the treatment, once it is approved, will be given to patients and will have an even greater variety of effects (since it is then applied to very different populations). Longino has written instructively about this effect in various domains of science:

The failure to attend fully to the interactions of the social group, including its females, in studying the males of the species has led to distorted accounts of the structure of animal societies. In toxicity studies, the focus on a single chemical’s toxic properties fails to inform us how its activities [are] modified, canceled or magnified by interaction with other elements in its natural environments. Focus on gene action has blinded us to the ways in which the genes must be activated by other elements in the cell. These models may well be empirically adequate in relation to data generated in laboratory experiments, but not in relation to potential data excluded by a particular experimental set-up.\textsuperscript{50}

Awareness of the effects of de-contextualization (or context-stripping) can do two things: it may suggest that we not limit our investigations to those that are de-contextualized and it may suggest that we need to work harder to ensure that when the results are used in practice the decontextualization is made very clear, as are the implications of decontextualization.

7.2.1.2. “Support the Healing Power of the Body”

Naturopaths are committed to supporting the healing power of the body. This draws upon the naturopathic belief in a vital force animating all living organisms. When CAM researchers play

\textsuperscript{50} Longino, “In Search of Feminist Epistemology,” 482.
by the rules of EBM, some individual alternative medical treatments will be proven efficacious. Treatments such as herbs (which resemble, and are testable like, drugs) are the best candidates for selective incorporation into mainstream medical care. This approach has also been referred to as the “greening” of mainstream medicine. Isolated alternative treatments are adopted by mainstream medicine while the underlying metaphysical view of health that originally produced the treatments is discarded. So, for example, though mainstream medicine might adopt particular herbs (St. John’s wort, ginseng, or garlic, for example), the naturopathic approach to health will likely be left behind.\textsuperscript{51} Though there are many similarities between mainstream and naturopathic approaches to health, including a first principle to “do no harm”, there are differences. The metaphysical account of health underpinning naturopathy emphasizes self-healing, highly individualized care, holism, vital force, and balance. This may be contrasted with the account of health currently taking over mainstream medicine as a result of EBM.

7.2.1.2.1. Health

Naturopaths raise concerns about the implicit, working definition of health adopted by mainstream evidence-based medical practitioners. This requires some background. In 1946 the World Health Organization (WHO) adopted the following definition of health: “Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.”\textsuperscript{52} The explicit recognition that health is not merely the absence of disease arose in response to widespread recognition that social and economic determinants of health (such as poverty) are significant in contributing to the global burden of disease and should be recognized and represented in any definition of health. Upon general acceptance, this new definition affected the structure of many international development plans developed in the decades to follow in which projects designed to provide adequate shelter, the basic elements of nutrition, and wells for clean drinking water in developing nations were seen as foundational to any health initiatives. This definition of health has not been amended since it was entered into force in 1948. After a

\textsuperscript{51} Similarly, while certain acupuncture points and procedures might be proven effective in RCTs and adopted into mainstream medicine, the underlying philosophy of traditional Chinese medicine, including the existence of the “qi” or vital force and the commitment to health as the balance of qi will be lost.

\textsuperscript{52} World Health Organization.
half century of use and wide acceptance by the global medical community, this definition is a good one (though still, of course, open to revision).\textsuperscript{53}

This all sounds good, from the perspective of naturopaths. However, there have been some important changes in the past half century, notably those brought about by EBM. If you narrow the research agenda to focus on a small range of research questions (i.e. those that can be studied in RCTs) and then require physicians to practice on the basis of this research, you are effectively narrowing the scope of treatments offered by (compliant) physicians to patients.\textsuperscript{54} As medical practice becomes more evidence-based, physicians will increasingly prescribe treatments to patients on the basis of a restricted evidential base. At the end of the day, this has implications for patients because they are offered a limited range of treatments. In addition, it is difficult for physicians to make operational the concept of health outlined by the WHO. Health becomes merely the state in which specific quantifiable symptoms or diseases are not present. How, after all, might a physician maintain a working commitment to the advancement of health as well-being (and not merely the treatment of disease) while consistently failing to care for patients with any environmental, psychological or socio-economic treatments and while paying little to no attention to the vast array of qualitative research designed to aid the physician in cultivating understanding and empathy for the patient’s subjective experience of illness?

Evidence-based physicians have been constrained to adopt a narrow working definition of health as a result of continued emphasis on a narrow scope of treatments. Medical care under the EBM approach is not advancing health in the WHO sense. Wendy Rogers writes,

EBM is a powerful influence upon the way we think about health and illness. Framing health problems in terms of the search for evidence of effective interventions tends to maintain discussions of health within the narrow biomedical model, diverting attention and resources away from alternative views.\textsuperscript{55}

\textsuperscript{53} It would be possible to focus instead on a definition of health provided by, for instance, the Canadian Medical Association, or the College of Physicians and Surgeons of Ontario. I prefer the WHO definition because it seems unlikely that any physician, anywhere in the world, would have a serious problem with it. I trade explicit commitment to the definition for broader application of my arguments.

\textsuperscript{54} This was discussed in Chapter Five. Also, the issue of compliance with evidence-based guidelines and standards is an interesting one. Many physicians do not comply with guidelines (hence the need for “knowledge transfer” initiatives and promotional work within hospitals). On this analysis, this resistance (however motivated) is at least partly a good thing (though, overall, it may still be bad) since it resists the narrowing of the scope of medical practice.

\textsuperscript{55} Rogers, “Evidence-based Medicine and Women,” 69.
Naturopaths are consistently reminded, through their principles of practice, of the vast array of complex human illnesses, which stem from sources as wide-ranging as the biological, social, economic, and environmental systems in which individuals are located, as well as the psychological, emotional, spiritual and physical states of the individual over time. Whatever narrow set of goals EBM might help us achieve, it is inappropriate as a tool for addressing all of these sources of illness and for providing targeted treatments for a wide range of diseases.

Standards of evidence, wherever they are designed and employed, serve to shape the direction of the field or discipline in which they are adopted. Standards of evidence that insufficiently recognize individual variation or which limit the type of question that can be asked in research to those which isolate simple, short-term relationships shape the understanding of health in the systems to which they are applied. This, in turn, limits the scope of “legitimate” treatments. This philosophical gap will only continue to grow as each of the specialized disciplines of mainstream medicine becomes more evidence-based. Acceptance of EBM will ultimately require naturopaths to commit to a metaphysical view of health that does not fit with their own. We are currently (however slowly and covertly) in the process of adopting increasingly narrow definitions of knowledge and health in medicine, and naturopaths have as much reason as any one else to remain committed to their current metaphysical account of health and to worry about the distorting impact of the adoption of EBM.

### 7.2.2. No Assimilation

The sameness approach has significant shortcomings for naturopaths and for researchers and practitioners in CAM more broadly. In particular, it leads to a decrease in attention to the sorts of principles and practices just discussed. This contributes to a decrease in methodological and theoretical diversity in medicine, which we know from CCE to be problematic. It is appealing, and no doubt tempting, for CAM researchers to play by the rules of EBM. It seems like an easy way to become scientific. However, if those rules are flawed, missing the chance to engage in critical debate about the merit of the rules and the ways they might be improved impoverishes all research. The diverse perspectives represented by CAM researchers will be lost if they choose to assimilate to EBM.
8. Building on Common Ground: Lessons and Future Directions

As noted earlier, alternative medical researchers are faced with three options. They can follow accepted EBM standards and design studies according to the demands of the evidence hierarchy (the sameness approach). Alternately, they can eschew these constraints in favor of studies that are ranked much lower on the evidence hierarchy such as qualitative studies, cohort studies and case studies (the difference approach). Finally, they can critically engage with the EBM standards, argue for the value of lower-ranked methods, and devise new research designs that more closely reflect the needs and goals of alternative medical practitioners. CAM researchers have raised a number of serious concerns with the standards of EBM. These concerns provide support for the suggestion that critically evaluating and re-conceptualizing EBM is not only the most reasonable and well-supported of the paths available to alternative researchers, but also that it is critical to the integrity of mainstream medical research. Since several of the principles of CAM research and practice are in tension with mainstream medicine, debate between the two serves to highlight the background assumptions of each. This is good for everyone.

In light of the arguments just presented, what CAM researchers might do is question the current standard of evidence to which they are being held: the evidence hierarchy designed by the EBM movement. Relevant questions would include, though by no means be limited to: Where did this standard of evidence come from? Is this standard best designed to answer all questions of medical significance? What are the assumptions underlying this approach to medical evidence? Does this epistemological view presuppose a particular metaphysical commitment regarding the nature of health and disease? This permits CAM researchers to avoid assimilation, while retaining an appreciation for empirical evaluation.

As we have seen in earlier chapters, many of the concerns with EBM encountered by CAM researchers are already fiercely debated within the conventional medical community. The extraordinarily influential EBM movement has not only rewritten the standards of evidence required of physicians and researchers, it has also galvanized debate within the medical community on, among other topics, the role of clinical expertise and case-studies in medicine and the challenges of finding appropriate evidence to inform clinical decisions at the level of the individual patient.
Concerns raised by alternative medicine that will likely resonate with mainstream physicians include the shift to physical/biological care at the expense of holistic care, the exclusion of social and environmental evidence from medicine through strict application of the evidence hierarchy, the devaluation of the patient in the decision-making process, the de-emphasis on qualitative evidence designed to help physicians understand the needs and values of patients, the lack of attention to pathophysiology and basic scientific research, and the narrow definition of health shaped by the evidence hierarchy. Several of these concerns have not so far been the focus of significant critical discussion among medical practitioners and researchers. Awareness of these critiques may provide conventional medical researchers with more reasons to join forces with alternative medical researchers and delve critically into the assumptions and theoretical goals of EBM. One of the possible implications for alternative medical researchers of choosing the third, more critical option is that they will be involved in the development of new and better research methods and outcome measures.

9. CAM Innovations and Use of Uncommon Methods, Outcome Measures, or Analysis

There have been a number of innovations that have come about because of the specific challenges faced by researchers in CAM. In addition, CAM research often draws attention to underused clinical outcomes measures, methods of data analysis, research methods, and frameworks for research programs. Each of these innovations or uses could have an impact on research in mainstream medicine. I will highlight a few innovations and a few of the underused methods CAM researchers make use of in order to illustrate the value of diverse perspectives in medical research.

9.1. Whole Systems Research

A group of international scholars have developed a framework for research into CAM therapies called Whole Systems Research (WSR). This framework was created to address the specific needs of researchers investigating whole systems approaches to health. Whole systems

---

56 Verhoef et al., “Complementary and alternative medicine whole systems research: Beyond identification of inadequacies of the RCT.”
approaches have three main characteristics: a unique theory of health and disease, complex interventions involving multiple modalities and combined effects, and highly individualized treatments. Naturopathic medicine, traditional Chinese medicine, ayurveda and homeopathy are all whole systems according to this definition. Naturopaths, as we have seen, subscribe to a theory of health and disease that emphasizes the self-healing powers of the body (the vital force), make use of as many as eight different modalities in the treatment of patients, believe that interventions have collective effects beyond their individual effects, and offer highly individualized treatments (often involving repeated consultations over long time periods).

Traditionally, medical research has been directed at the evaluation of specific components of whole systems approaches, such as particular herbs or styles of massage. While this research is useful at answering questions about the efficacy of particular treatments in isolation, it does not offer any sense of the effectiveness of such treatments as they are used in conjunction with other treatments, and within the context of a strong practitioner-patient relationship. As such, these results fail to investigate the sorts of questions that matter to CAM practitioners and their patients. There is a practical interest expressed by CAM practitioners and patients in whether something works in practice, above and beyond any questions about specific efficacy. These whole systems questions, argue proponents, are worthy of investigation in their own right. This in no way implies that research into particular components is unnecessary, only that it is insufficient.

The proposed framework for research into whole systems “targets the study of complex CAM therapies as system-level phenomena, as opposed to single-agent or uni-dimensional effects.” In other words, whole systems research takes seriously the need for evaluation of the effectiveness of whole systems rather than insisting only on research that reduces the evaluation of whole systems to their component parts. It recognizes the commitment of CAM practitioners to the inseparability of individual components of health care, the importance of the practitioner-patient relationship, and the context-specificity of results. In short, it recognizes the shortcomings of context-stripping in research, and values contextual research. Whole systems research (WSR) is designed to contribute to research on the effectiveness of whole systems approaches to health.

57 Recall that efficacy indicates how well a treatment works in idealized settings while effectiveness indicates how well a treatment works in regular practice, with typical patients.
58 Ritenbaugh et al., “Whole systems research,” 32.
The aim of WSR is to use appropriate research designs and methods so that all aspects of any internally consistent approach to treatment, or a whole system, can be assessed within its unique explanatory model.\textsuperscript{59}

The novel contribution of the WSR approach is its emphasis on model validity. Medical researchers have traditionally recognized the importance of the internal validity (or efficacy) and external validity (or effectiveness) of the results of research. Model validity refers to two further elements: 1) the fit between the research method and the theory of health and disease underlying the therapeutic intervention under investigation 2) the fit between the research method and the standard therapeutic context. Internal, external, and model validity are all interrelated, but model validity is often taken for granted. Mainstream medical researchers tend to presuppose theories of health and disease that are mechanistic and reductionistic: researchers aim to isolate causes and effects along physiological pathways. As a result, research methodologies, such as the RCT, have been developed in order to aid in the careful isolation of causal relationships between treatments and their pre-specified effects. WSR advocates do not argue that such research is flawed. What they aim to do is draw attention to the ways in which assumptions about health and disease and the best way to understand phenomena fit with certain research methodologies more so than with others.

Researchers in mainstream medicine have been able to ignore issues of model validity because they share the same assumptions about the nature of health and the physician-patient relationship. CAM researchers – particularly those researching whole systems approaches to health – encounter challenges when faced with standards of evidence that are incongruous with their theories of health and disease or methods of practice. So model validity is a general criterion for research, but is encountered as a particular problem primarily in CAM research. Model validity can be enhanced in a number of ways.

9.1.1. Double Classification

Consider the following case. A researcher is interested in evaluating the effectiveness of traditional Chinese medicine (TCM) in patients with hypertension. She proposes an RCT which

\textsuperscript{59} Verhoef et al., “Complementary and alternative medicine whole systems research,” 207.
enrolls patients with high blood pressure and then divides them into groups receiving TCM and standard mainstream care. Results indicate whether, as a group, the patients receiving TCM did better or worse than those receiving standard care. The problem with this trial is not just that it fails to be representative of the context of medical practice and thus has weak external validity (a shortcoming common to RCTs) but that it actively misrepresents TCM and, as a result, fails to evaluate anything meaningful at all. Such a trial has very weak model validity. It ignores the commitment to individualized care at the core of TCM practice and lumps together patients under mainstream medical diagnoses that would not have been grouped together under TCM diagnoses. According to TCM, a mainstream medical diagnosis of hypertension might be anything from “rebellious qi in the liver and stomach” to “blood (yin) deficiency” to “liver and kidney yin deficiency” depending on the particular individual. From the perspective of a TCM practitioner, then, there is no homogeneity – no common condition – that brings these particular trial participants together. This problem is compounded when TCM practitioners are asked to produce “the treatment” for hypertension, so that it can then be standardized and given to all patients in the TCM treatment group. At this point, inappropriate diagnosis has been compounded with treatments that no TCM practitioner would necessarily offer to the particular trial participants. While this second problem can be avoided by offering individualized TCM treatments within the TCM group, the first problem endures. At the end of the trial, the results are meant to provide an indication of the efficacy of TCM for patients suffering from hypertension. This is not necessarily a meaningful question from the perspective of the TCM practitioner for the reasons just outlined.

In order to address this problem, advocates of WSR have suggested that research subjects in these sorts of trials should be diagnosed by mainstream medical physicians as well as practitioners from the whole system under investigation. This double classification then leads to complex multiple-stream trail designs, and in turn greater model validity. It also makes the trial very difficult and the statistical analysis challenging. However, the ability to ask and answer questions of value to both TCM and mainstream researchers makes these double-classification schemes useful and popular.

---

60 These diagnoses do not refer to physiological organs (kidney, spleen) but to qi meridians that have no physiological corollary. Possible diagnoses from: Maclean and Lyttleton, Clinical Handbook of Internal Medicine Volume I.

61 Double classification was originally proposed by Fisher and colleagues in their article, “Effect of homoeopathic treatment on fibrositis (primary fibromyalgia).”
9.1.2. Variations of the RCT

CAM researchers have drawn attention to neglected or unpopular variations of the RCT. I will discuss some of these below.

9.1.2.1. Pragmatic Trials

Pragmatic RCTs provide results that can be used immediately in practice. Rather than evaluating the efficacy of a treatment under ideal conditions, such trials evaluate the effectiveness of a treatment under usual conditions. Pragmatic RCTs are gaining popularity in both mainstream and alternative medical communities. Traditionally, RCTs have been designed to ensure the internal validity of results, but this has been secured at the expense of external validity. This is a result of the trade-off between specificity and generality (or accuracy and scope) of claims. This is most often achieved by using strict inclusion and exclusion criteria for trial participants. As a result, trial participants often differ significantly from those who later receive the treatment from their physician or practitioner. The more homogenous the research group, the more researchers can aim for internal validity. However, it is external validity that matters most to physicians and practitioner, who must determine whether the results of an RCT can be applied to the case of a particular patient.

Pragmatic trials take care to represent “real world” patients, diseases, and health-care settings. Heterogenous groups of patients are more common in these trials, and as a result trial participants more closely resemble the target treatment group in the greater population. So, for instance, if the treatment is targeted to elderly women, then it is elderly women on whom it is tested, even if that means including patients with comorbid conditions. Care is taken to minimize the exclusion criteria for the trial. Pragmatic trials also mimic the complexity of regular patient care by allowing a greater degree of flexibility in offered treatments. In the evaluation of naturopathy, for instance, a pragmatic trial would allow different naturopaths to make use of different modalities in treatment depending on their assessment of the individual needs of the patient. So while one patient might receive acupuncture, the next might receive mostly nutritional treatment. This more closely reflects what would happen in practice, and allows
researchers to acknowledge their interest in answering the more pragmatic question. The pragmatic RCT allows for black box evaluation, in that there is flexibility in the precise treatment offered – within certain limits. Some proponents have also argued that the shift away from placebo-controlled trials to standard care controls is an innovation of the pragmatic trial.62

9.1.2.2. Factorial Trials

Factorial designs “compare single modalities to a combination of modalities to allow for the assessment of multiple interaction effects between treatments.”63 Such trials have multiple treatment groups, each with different combinations of treatments. It is still possible to isolate the effectiveness of particular treatments when compared with other combinations. For instance, in the U.K. back pain exercise and manipulation (BEAM) trial patients received standard care plus one of the following: “(1) an exercise program; (2) manipulation; or (3) manipulation followed by exercise.”64 This allowed researchers to compare manipulation and exercise to each other, and, significantly for CAM researchers, to examine the potential cumulative effects of the two treatments against each one isolated from the other. The ability of this method to isolate data on efficacy, while also investigating cumulative effects of many interventions, makes it well suited to upholding model validity and therefore appealing to CAM researchers.

9.1.2.3. Preference Trials

In a preference RCT patients are asked whether they have a preference for one of the treatments offered in an arm of the trial and if they have a preference they are assigned to that group. If they do not, they are randomized as usual. This allows for analysis of randomized patients as usual, but also for analysis of the efficacy of different treatments by those who chose their treatment, and analysis of the efficacy of a specific treatment in those randomized to it compared with those who chose it. Patients who seek out CAM therapies often have strong opinions about which

63 Verhoef et al., “Complementary and alternative medicine whole systems research,” 208.
64 Ibid.; UK BEAM Trial Team, 1377.
medical treatments they receive. As a result, they are more likely to opt out of RCTs when they are informed of the chance that they will not receive the treatment they want. This type of design allows researchers to keep such patients in the trials. Because CAM practitioners often have a deep commitment to patient-centered care and patient choice, this design is also a better fit from their perspective since it is more respectful of patient preferences.

9.1.2.4. N-of-1 Trials

RCTs produce data that is averaged over the patients in the trial. Physicians and practitioners encounter individual patients. The gap between the average patient (after inclusion and exclusion criteria) and the individual patient is a significant one, and is the first thing critics of RCTs mention when listing the problems with the RCT. The data produced by a n-of-1 trial is highly individualized and as such is particularly well-suited to the evaluation of CAM interventions. It also represents an improvement over case-studies as the individualized method of choice for evaluating CAM interventions. The n-of-1 is most useful in cases where the condition is persistent and stable, and the patient is highly motivated. These are quite often the health status and characteristics of CAM patients, who tend to have chronic health problems and have taken the initiative to seek out CAM care (often at their own expense). The gains made in individualizing data in the n-of-1 trials are arguably lost in generalizability, but that may be a price CAM researchers and practitioners are willing to pay. Also, in light of arguments suggesting that traditional RCTs are not that generalizable either, this may just be a case of recognizing a generalizability problem in medical research (which could just turn out to be a function of the type of research it is and the types of idiosyncractic, complex and social subjects it necessarily investigates). Finally, the results of individual n-of-1 trials can be brought together and meta-analyzed if more generalizable data is desired (though this may suffer from the same problems meta-analyses tend to suffer from).

---

65 Fitter and Thomas, “Evaluating complementary therapies for use in the National Health Service.”
66 Johnston and Mills, “N-of-1 Randomized Controlled Trials.”
67 Astin.
9.1.3. Appropriate Outcome Measures: MYMOP

CAM researchers have developed a new outcome measure for research and practice. The Measure Yourself Medical Outcome Profile (MYMOP) was developed by U.K. medical researcher Charlotte Paterson and first presented to the medical community in the *British Medical Journal (BMJ)* in 1996. As the name indicates, the questionnaire is designed to track changes in outcomes that matter to patients. This is in contrast with outcomes questionnaires that presuppose the concerns identified as most important by physicians or that are based on preset quality of life indicators. Patients initially record one or two symptoms that concern them most (usually the main reasons they are seeking help) and one activity of daily living that has been restricted by these symptoms. They assign a value to these concerns on a seven point scale, where seven is “as good as it could be”/“able to do it normally”. They also respond to one question about their overall well-being. Later, the patients complete follow-up questionnaires in which the original concerns are retained and they assign new values on the seven-point scale. There is also an opportunity at this stage for patients to add one new concern and rate it. The questionnaire measures improvement or deterioration in the concerns identified by patients. Originally designed as a tool for primary care, the questionnaire is also well suited to be used as an outcomes measure in research. The questionnaire has been validated according to standard validation measures.

Paterson’s *BMJ* article is written in a way that downplays the connection between the development of the MYMOP and CAM. In spite of appearances, and though Paterson is trained as a medical doctor, she has a long affiliation with the CAM community and makes use of complementary therapies in her practice. It is this appreciation for CAM and frustration with usual objective outcome measures (such as the SF-36) that motivated Paterson to create the MYMOP questionnaire. Quality of life measures are relatively crude and often fail to capture changes that are important patients. As Paterson puts it, in a response to a critique of her article, how else might a practitioner capture the patient’s concern with a symptom described as “a

---

69 The most recent version of the MYMOP questionnaire, along with guidelines and advice, is available to the public at [http://www.bristol.ac.uk/hsrc/research/other/mymop/index.html](http://www.bristol.ac.uk/hsrc/research/other/mymop/index.html).

70 Paterson, “Measuring outcomes in primary care.”

71 For details on the validation process see the Paterson article (Ibid.).
feeling of water gushing over the head” or “a burning sickness”?

Functional outcomes such as the ability to return to work can also be highlighted by the patient. The questionnaire was created in order to improve the ability of practitioners to measure changes in patient-centered outcomes and to improve patient-centered care. The principles of patient-centered care are foundational to most CAM practices.

In the original *BMJ* article Paterson included one section specifically on CAM. In it, she wrote, “In the evaluation of complementary therapies a patient generated measure may overcome the problem of the different diagnostic frameworks of different disciplines.” The questionnaire will be helpful in cases where, for instance, diagnostic differences exist between traditional Chinese medicine (TCM) and mainstream medicine (as discussed above). Researchers can use the MYMOP questionnaire to bypass some of these differences. As Paterson points out, this has implications for mainstream medicine as well, since physicians often disagree about appropriate diagnoses. Because the questionnaire allows patients to identify their most significant concerns, it sets outcomes independent of particular diagnoses. In addition, the inclusion of CAM practitioners in the validation of the questionnaire and extensive use of MYMOP questionnaires within the CAM community since it has been made available, as well as the development of more recent variations including the Measure Yourself Concerns and Wellbeing (MYCaW) questionnaire developed specifically for complementary cancer therapies provide some indication of the ways in which CAM research has shaped the development and evolution of the MYMOP questionnaire. Finally, in the management of chronic conditions, the primary concerns are with the alleviation of symptoms and prevention of (further) diseases. CAM

---

72 Paterson, “Author’s Reply,” 627.
74 Paterson cites a study on the differences in diagnosis found among general practitioners dealing with upper respiratory illnesses: Howie, “Research in general practice.”
75 There are, of course, potential problems with such a measure, including the possibility that patients are radically mistaken about what they should be most concerned about. Imagine, for instance, a patient who is most concerned with her acne when she is dying of pancreatic cancer and has only a few weeks to live. Patients are not always the best judge of outcomes that matter to their health. At the same time, it is precisely this shift to patient-centered outcomes and measures that, for better or worse, has become the new ideology of medical care (at least in Canada). Insofar as this is the case (I am not going to delve any further into this debate here), this questionnaire is undoubtedly useful. Another similarly useful outcome measure is the ‘Goal Attainment Scaling’ (GAS) questionnaire proposed by Becker et al. in their article “Goal Attainment Scaling to Measure Individual Change in Intervention Studies.
76 For further information on the MYCaW questionnaire see Paterson et al. “Measure Yourself Concerns and Wellbeing (MYCaW).” The MYCaW questionnaire is publicly available at [http://www.bristol.ac.uk/lsrc/research/other/mymop/mycaw/index.html](http://www.bristol.ac.uk/lsrc/research/other/mymop/mycaw/index.html). For an example of the MYMOP used in CAM research see Chapman et al., “A descriptive outcome study of 291 acupuncture patients.”
therapies are frequently used by patients suffering from precisely these sorts of long-term, chronic illnesses. The emphasis on therapeutic regimens that seek to alleviate patient-defined concerns seems particularly well-suited to chronic illnesses in which a cure is not possible.77

9.1.4. Data Analysis

Some of the tools of data analysis, borrowed from the social sciences, have also been suggested as ways of getting at some of the individual patient data hidden within the results of RCTs. These methods would include structural equation modeling, path analysis, and event stream analysis, and would help in identifying some of the individual reactions to treatments that often get lost in the “average patient” data of RCTs. These analytic tools are of significant interest to CAM researchers for many of the reasons already outlined.

9.1.5. Observational Research

RCTs are notoriously short. The average RCT is 6 weeks long. While it is possible, in principle, to carry out RCTs for long periods of time it is prohibitively expensive. As a result, it tends to be observational studies that investigate treatments over longer periods of time (years and even decades). Because the success of modern (mainstream) medicine has largely come about in acute care, CAM practitioners quite often end up with patients who have chronic problems, from back pain to asthma to chronic fatigue. It is the long-term observational research that is more likely to get at information about how patients with chronic conditions do over a long period of time on a particular intervention. There is a nice fit between CAM and observational research. Carefully designed, rigorous observational research should be a focus for CAM researchers.

9.1.6. Qualitative Research

CAM researchers are committed to mixed methods approaches and assign equal value to quantitative and qualitative research methods. CAM researchers argue that each set of methods

---

77 The Arizona Integrative Outcomes Scale is another example of an outcome measure developed by CAM researchers. Bell et al., “Development and validation of a new global well-being outcomes rating scale for integrative medicine research.”
helps us to understand certain aspects of complex interventions and we are best off when we have as many different descriptions of phenomena as possible. It is also qualitative evidence that, as noted earlier, answers the ‘why’ questions about patient behavior. Because patient compliance is so important in CAM (since so many interventions, for example nutritional changes, must be maintained and monitored by the patient over long periods of time) understanding the reasons why patients act the way they do is very important to CAM practitioners.

10. Lessons from CAM

These sorts of innovations and defences of underused methods could be an important contribution to a shift in the conventional understanding of best evidence that accepts the need for, and value of, rigorous testing in medicine without unduly and unnecessarily restricting the scope of evidence. The work being done in alternative medicine right now is not merely something to be dismissed as fringe science, but is an attempt to re-work the EBM hierarchy for the benefit of patients in a way that is consistent with values underlying all of medical practice, broadly construed. This is a movement that all practitioners, whether alternative or mainstream, have good reason to welcome and encourage.

New methodologies and approaches to solving problems can be invented at any time. These innovations are driven in part by the need for methodologies that can answer new questions raised as science progresses and social contexts change. Given the ever-changing nature of the questions and problems encountered by scientists in a variety of fields, scientific communities should not limit their standards in such a way as to exclude the possibility that new and better methodologies might come along at any moment. Similarly, as the questions under investigation change over time, certain methodologies may come into fashion. For these reasons, the cognitive goals and practical ends of any community will not be met best by a static commitment to particular methodological approaches or regulatory standards, but rather by an open-ended commitment to certain community standards.78

Scientific communities are held together by shared community standards, and it is when these standards are met that regulatory standards can be designed. Regulatory standards are not static, but must be responsive to the changing cognitive needs and practical ends of the

78 The distinction between regulatory and community standards is offered in Chapter Two.
community. As we saw earlier, advocates of the EBM hierarchy require those outside the mainstream medical research community to conform to existing regulatory standards. It might have appeared, unreflectively, as though these regulatory standards provide boundary conditions on who should have a voice in the medical research community. However, if regulatory standards are imposed on others, diversity is compromised. Stabilizing scientific inquiry in this way demonstrates a dogmatic commitment to particular goals and ends, and a belief that one has identified the important questions that need to be answered within a domain. Those who have other questions, or other goals/ends, are seen as deviant rather than as presenting an opportunity for change. But, as we saw in Chapter Two, CCE does not suggest that regulatory standards can be imposed on communities. In order for outsiders to be included in a community debate, there must be some overlap in community standards (shared public standards) and the domain of investigation.

What this analysis demonstrates is the inadequacy of a response to CAM which simply requires that CAM researchers completely assimilate and adopt all community and regulatory standards set by the mainstream. It is not necessary, as we saw in Chapter Two, for members of non-mainstream communities to share all standards with mainstream communities if they hope to have a voice in critical debates. The requirement is, rather, that they share at least one of the community standards, which in turn follows from a common cognitive goal or practical end within a shared domain. A clear commitment to empirical adequacy serves as this point of overlap between at least some CAM research communities and the mainstream medical community. In what follows, I consider one likely criticism of my argument thus far. This critique arises out of a tension between the requirement to cultivate diverse perspectives and the need to maintain rigorous standards for medical research.

11. Critical Analysis

To sum up: diversity is good for medical inquiry because it contributes to the objectivity of medical research. EBM suppresses diverse research methodologies through the evidence hierarchy, and widespread social acceptance of EBM has led to increased pressure on researchers from a variety of theoretical backgrounds to assimilate into the EBM model. Based on the
arguments offered in Chapter Two, we have good reason to consider the negative implications of this decrease in diversity for the future of medical research.

11.1. Objection

One anticipated concern with this argument stems from the recognition that science proceeds not only through the advancement of new ideas, but also by slowly and tediously working out the details of accepted theories through experimentation. How, if we become too interested in cultivating diversity, are we going to avoid endless testing and retesting, endless new proposals and endless new ideas?\textsuperscript{79} Progress seems to require some sort of balance between what Longino identifies loosely as the two missions of science. These are: the knowledge-extending mission, which involves the construction of comprehensive accounts of the natural world through piecemeal puzzle-solving and the gradual extension of theories to new data, and the critical mission which aims at an accurate (or true) description of the world through extensive testing, retesting, rejecting and reformulating of hypotheses.\textsuperscript{80} If we allow a great number of diverse perspectives in science we will spend all of our time testing and rejecting new hypotheses instead of advancing and perfecting already established theories.

11.2. Reply

I suggest that the EBM hierarchy is a straightforward instantiation of the knowledge-extending mission of science. Longino describes what happens when scientific inquiry tends more toward this sort of conservative approach and subordinates concerns with its critical mission. I will quote her description at length:

The systematic and unifying treatment of phenomena enables us to interact with the natural world with reliable expectations. A methodology that legitimizes the stabilization of inquiry thus serves some constitutive ends of knowledge seeking. It must also, however, subordinate science’s critical function in order to avoid the endless testing and constant generation of new explanatory frameworks that would subvert knowledge extension, and it must disguise that subordination to deflect the accusation that the

\textsuperscript{79} Longino, \textit{Science as Social Knowledge}, 223.
\textsuperscript{80} Ibid., 32-34.
sciences are not after all concerned with truth. One way to achieve this disguise is through the adoption of an account that minimizes the need for and role of criticism beyond hypothesis testing, that is, by an account that can render invisible the role of background assumptions.81

This describes the progress of the EBM movement with surprising accuracy. EBM was created as an anti-authoritarian, even liberatory, movement. Proponents of the earliest version of EBM explicitly called for the need to supplant tradition and authority, and interject objective evidence in its place.82 The EBM approach is positioned as a critical approach that emphasizes active, engaged critical thinking. But this is not the practice – at least as represented in medical journals – as the structure of EBM means that critical debate tends to be restricted to asking how sources of evidence such as intuition and experience can be fit into the evidence hierarchy or added on to the results of the hierarchy. Questions about the assumptions underlying the hierarchy itself are rarely seriously engaged.83 According to Longino, emphasis on the conservative, knowledge-extending mission of science “requires that its critical mission be blocked.”84 Overemphasis on the knowledge-extending mission of science often “discourages the investigation of alternative frameworks.”85 In fact, “those who might otherwise be inclined to do so [investigate alternative frameworks], in the spirit of free inquiry, are dissuaded by a combination of related phenomena: from the desire to be a ‘good’, effectively orthodox scientist to the lack of attention accorded to nonmainstream ideas.”86 I hope to have demonstrated the ways in which EBM is doing just this.87

Those who are concerned that encouraging the cultivation of diverse perspectives will stall the production of good research and waste time and resources are right to point out the need for balance in the two missions of science. An overemphasis on diversity is not beneficial to scientific inquiry. Of course, a diminished emphasis on diversity is similarly bad for scientific

81 Ibid., 224.
82 Interestingly, commentators have noted that EBM has led to the creation of new forms of authority (Cochrane collaborations, guidelines, and the evidence hierarchy) and is now most often identified as an approach that preserves, rather than overturns, the status quo.
83 Notable exceptions to this trend do exist. See in particular the work by Upshur, Tonelli and Tanenbaum, and publications in the annual issue of the Journal of Clinical Evaluation dedicated to critiques of EBM.
84 Longino, Science as Social Knowledge, 223.
85 Ibid., 225.
86 Ibid.
87 As an interesting side note, Longino also points out the ways in which emphasis on the knowledge-extending mission of science serves to shore up the professional boundaries of particular scientific communities: “struggles to exclude would-be scientists from the professional scientific community…are, among other things, battles to reduce the number of formative assumptions, thus stabilizing the object of inquiry and enabling the development of theory under a unifying or unified and eventually transparent set of values.”87
inquiry. I hope to have offered some reason to believe that the danger in medicine is not currently one of embracing too much diversity. In addition to my arguments on this point, former editor of the *New England Journal of Medicine*, Marcia Angell, has argued that, at least in North America, medical research is plagued by a lack of innovation, medical practice is circumscribed by strict rules and guidelines and available medical therapies exhibit an increasingly narrow focus on acute conditions and pharmaceutical treatments.\textsuperscript{88} The trend in each of these areas is one away from diversity. If EBM and other social forces have shifted the balance of research in one direction, research that stimulates a return to equilibrium can only benefit medicine. I suggest that contemporary EBM proponents should be encouraged to take a more diversity-friendly stance, or at least to consider (and publicly debate) the need for such a stance. This might include a reconsideration of the central role of the evidence hierarchy in medical research and practice, a return to the emphasis on critical discussion that was originally at the core of the EBM, the extension of this critical attitude EBM’s own background assumptions, and a re-evaluation of the demands made of non-mainstream researchers. This will not mean the rejection of empirical standards or dissolution of the boundaries between good and bad science.

Medical researchers must maintain a commitment to knowledge extension and to critical evaluation. I argue that the EBM movement has not been sensitive enough to the importance of this balancing act. Further, I suggest that the conservatism underlying the present imbalance has significant practical and philosophical implications. These implications warrant open discussion by the members of the medical community, broadly construed.

12. Nursing

Finding common ground in the EBM debate is critical, given that, as Verhoef and colleagues remind us, the issues raised apply also to complex interventions in mainstream medicine, such as “multidisciplinary chronic care, patient-centred primary care, psychotherapy, and palliative care.”\textsuperscript{89} In fact, there is a great deal of common ground between the sorts of critiques of EBM made in these other areas and those made in CAM contexts. I will mention a few of the comments made by nurses as an example. Like CAM researchers and practitioners, nurses are

\textsuperscript{88} Angell, *The Truth about the Drug Companies*.
\textsuperscript{89} Verhoef, et al. “Complementary and alternative medicine whole systems research,” 211.
under a great deal of pressure to provide evidence of the effectiveness of their interventions and care. Recently, researchers writing on EBM have made the following claims:

Currently, traditional practices retain a stronghold in nursing and health care. Unless the initiative is taken to develop and search for evidence to uphold or refute these practices, nursing will not be in a position to answer, with confidence, questions about why, or if, these traditional practices ought to be maintained or changed.\textsuperscript{90}

It is imperative that the nursing profession continue to explore the philosophical perspectives that underscore evidence-based practice and their implications for decision-making in nursing practice.\textsuperscript{91}

There are similarities between the sorts of challenges raised by CAM researchers and those raised by nurses. For instance, in order to prevent nursing from becoming dehumanized or decontextualized, nurses suggest that qualitative research should complement quantitative research. A clear emphasis on the context-specificity of nursing is said to be vital to connecting nursing research and practice.\textsuperscript{92} Critiques of EBM come from such diverse perspectives as: CAM researchers, nurses, psychiatrists, palliative care specialists, surgeons, public health officers, and administrators. While I have highlighted the concerns raised by CAM researchers, attention to the diverse perspectives offered in each of these areas would only strengthen the critique of EBM.

I argue that diversity is critical to the advancement of knowledge in medicine. This is because communities expressing strong dissent from mainstream views are able to identify and draw attention to methodological and theoretical assumptions underlying research programs within the mainstream community. One advantage of this enforced self-reflection is that it enables members of both mainstream and alternative communities to correct flaws in their systemic assumptions. In addition, it promotes greater understanding of the epistemological foundations of standards of good evidence and requires that researchers justify such standards against alternatives. Unfortunately, as I demonstrate in the above, the recent EBM movement constrains the cultivation of diversity in medical research. It does so by suppressing diverse research methodologies through the evidence hierarchy, and by increasing the external social pressure to adhere to these standards. I argue that EBM advances the knowledge-extending

\textsuperscript{90} Romyn et al., 187.
\textsuperscript{91} Ibid., 184.
\textsuperscript{92} Ibid., 184-8.
mission of medical research at the expense of the critical mission, and suggest that this imbalance is dangerous. Attention to the ways in which EBM suppresses diversity will contribute to a more appropriate balance for medical research.

13. Summary

In light of the epistemological arguments for diversity provided by CCE and the ways in which EBM suppresses methodological and theoretical diversity, there are good reasons to be concerned about the EBM movement. Further, attention to diverse methodologies and theoretical frameworks can be helpful in exposing the background assumptions of mainstream research. This is a significant accomplishment according to the CCE approach to knowledge-production. It would be astonishing to find that the standards currently adopted by mainstream medicine are perfect, and that we can incorporate entirely different systems of healing into the mainstream system without having to make any changes or modifications, especially given the immature state of our current standards (EBM is just 15 years old, after all). If alternative medical researchers concede to the demands of the EBM hierarchy and mainstream medicine slowly integrates elements of alternative medicine into common practice, there may not be any change to the current hierarchy. That would be a shame, because as I hope to have established in this discussion, there is significant common ground between the concerns with EBM raised by alternative medical researchers and those identified by mainstream medicine. When the diverse communities of medical practitioners and researchers are all engaged in critical discussion we are in the best possible position to identify and create truly excellent standards of evidence. Conventional medical practitioners and researchers, and certainly patients, have much to gain from attention to diverse perspectives on health – and much to lose if alternative medical researchers submit uncritically to the hegemony of EBM.
Evidence-based medicine aims to increase the objectivity of medical research and practice. Objectivity is regularly used to signify “everything from empirical reliability to procedural correctness to emotional detachment.”¹ It is not immediately evident what people mean when they speak about the objectivity of research results, outcomes, methodologies and individuals – particularly when they speak of these all in the same breath. Philosopher of science, Heather Douglas, has recently identified eight distinct senses of objectivity in common use in the analytic philosophical literature.² In the first part of this chapter, I outline these senses of objectivity and explain the strengths and weakness of each sense. Following this, I draw upon this careful catalogue in order to characterize the sense of objectivity underlying my version of critical contextual empiricism (CCE). I suggest that CCE upholds what Douglas calls interactive objectivity. CCE eschews any expectation that individuals remain detached or neutral in the course of inquiry. Instead CCE emphasizes the public nature of knowledge and identifies features of communities that contribute to the transformative criticism of proposed hypotheses. Thus, according to CCE, interactive objectivity is secured through discussion and debate within a diverse community.³

In the latter sections of the chapter, I discuss the importance of these types of community discussions for the standards of evidence proposed by EBM. I argue that there has been an overriding fixation on procedural objectivity as a result of EBM. Drawing on Douglas’ catalogue and the CCE account of interactive objectivity, I identify potential dangers with this fixation. In particular, I raise concerns about the inability of such an account to address the influence of unidentified social values on medical research. According to CCE, however good procedural objectivity may be, it is not sufficient for robust objectivity. In this chapter, my aim is twofold. First, I aim to provide a comprehensive description of the CCE account of objectivity and relate

---

¹ Daston and Galison, “The Image of Objectivity,” 82.
³ The modified version of CCE outlined in Chapter Two specifies what is meant by diversity.
it to standard accounts of objectivity. Second, I aim to demonstrate, using the medical research context, how and why this particular understanding of objectivity is valuable. In pursuing this second project, I provide arguments about the shortcomings of EBM and the ways in which the medical community can improve the interactive objectivity of their standards of evidence. Drawing on the modified CCE norms, I outline the sorts of interventions in medical research and practice required by interactive objectivity, including: elimination of confidentiality agreements signed by researchers, disclosure of competing interests in publications, prohibition on ghost authorships, public registration of research, open peer review, open access to publications, open submission processes, increased public funding of research, less reliance on guidelines, greater awareness of the collective nature of medical decisions and, most importantly, greater analytic training for medical professionals. I argue that the narrow focus on procedural objectivity currently adopted in medicine must be expanded to include support for the social institutions that enable conditions of interactive discussion and debate.

1. A Short History of Objectivity in Science

Within contemporary analytic epistemology and philosophy of science a consensus on the appropriate way to define objectivity is, as Elizabeth Lloyd puts it, “glaring in its absence.” She continues, “[T]he emphasis on the limitations of objectivity, in specific guises and networks, has been a continuing theme of contemporary analytic philosophy for the past few decades.” Contemporary epistemologists hold a wide variety of views on the nature of objectivity. Arthur Fine has described the concept as a “hodgepodge” of ideas and ideals. If we are to talk intelligibly about the nature of objectivity, we must begin by acknowledging that objectivity is a complex concept and that its definition has been disputed from a variety of metaphysical and epistemological perspectives throughout the history of its use. This is true even when we narrow our focus to the realm of science. As historians Lorraine Daston and Peter Galison point out, the concept has evolved over time in response to the concerns of scientists in different time periods. According to Daston and Galison, we label something objective when it is opposed to that which

---

5 Ibid., 375.
is considered the most dangerous form of subjectivity by people of the day. While scientists throughout history have endeavored to free inquiry from the negative influence of subjective values, the perceived sources of those values have fluctuated over time.

It is difficult to speak of the sciences broadly, since the history of specific sciences have followed very different paths, but Daston and Galison draw some general lessons from botany about the development of our current concept of objectivity. Their analysis begins in the eighteenth century, when scientists were primarily concerned with objects that were idiosyncratic and difficult to capture. In order to correct for this, there was an emphasis on the training and expertise of scientists and the professional character of objectivity. The common belief was that disciplined geniuses or sages were needed to get at what is real in the unruly world because they are able to see beyond mere appearance to the ideal form of objects. Thus, expert interpretation was an integral part of objectivity. In the nineteenth century, concern arose about the inner temptations of scientists. The greatest obstacle to objectivity became the impure influence of the scientists on the objects of study. Thus, the focus was on heroic self-restraint and the personal (and individual) character of objectivity. Scientists were expected to remain pure and let nature speak to them. In the nineteenth century, it was the scientist (not the world) that was unruly and in need of control.

The early twentieth century marked a brief return to the celebration of unconscious skills and interpretation with a renewed focus on expert judgement. In addition, when scientists in the twentieth century realized that heroic self-restraint could not fully eliminate values, their attention shifted to social procedures designed to eliminate these individual idiosyncrasies. So, according to Daston and Galison we see the rise of a strict scientific method in many areas of science, and though in some areas there may be a lingering commitment to interpretive expertise, this is largely overridden by adherence to ostensibly objective procedures. That is to say, we are currently concerned more with the unruly scientist than the unruly world. And we do not think the unruly scientist has the self-restraint to deal with the problem on his or her own, so many scientific communities have adopted rules and guidelines of practice for researchers in the hopes of constraining the ability of researchers to bias research. The concept of objectivity is complex.


---

7 Daston and Galison, “The Image of Objectivity.”
and evolving, and though it is generally opposed to subjectivity, the perceived sources of
dangerous subjectivity have shifted over time.⁸

2. Catalogue of Objectivity

In order to isolate the elements of objectivity that have received varying degrees of attention
from scientists over the years, I turn to the recent work of Heather Douglas. As mentioned above,
Heather Douglas has identified eight distinct senses of objectivity in common use in the
philosophy of science.⁹ This catalogue will be the starting point for my investigation, as it helps
to disambiguate the various bundles and strands of objectivity in contemporary use. In the course
of putting this catalogue of objectivity together, Douglas avoided having her work overshadowed
by the realism/anti-realism debate by focusing her analysis on the different sorts of objectivity
inherent in knowledge-productive processes. Limiting discussion to objective processes allows
Douglas to avoid taking a position on whether the results of such processes are ‘really real’,
‘reliably produced’ or merely ‘useful’. This focus on objective processes is especially well suited
to understanding feminist accounts generally because they tend to focus on practices and
processes. In addition, the approach is prudent since epistemological accounts that rely
extensively on metaphysical accounts are required to take on the onerous responsibility of
defending both.¹⁰ This careful catalogue of senses of objectivity is significant because it helps us
to identify common bundles of these senses adopted in various realms of science. Some of the
senses are more prevalent than others, and, as I will suggest, some might be better catalogued
with a different family of terms. The catalogue is, nonetheless, a helpful starting point for an
analysis of objectivity.

⁸ As Ian Hacking eloquently puts it: “Objectivity is not the less massive, impenetrable, resistant, because it is the
product of our history. But when we get close enough to run our hands across the rock – or rather, conglomerate –
we shall feel its fissures and notice how different is its texture from that smooth surface that we seem to observe
from afar, before we attend to the innumerable details that are its only origin and which constitute its substance.”
¹⁰ Longino, Science as Social Knowledge. 17-19. See also Longino, The Fate of Knowledge, 140, where she writes
“The kind of social epistemology I advocate does not commit itself to any metaphysics of nature… [P]hilosophers
may use metaphysical assumptions implicitly in their development of a general epistemology. This is, as I have
argued, a flaw in such accounts, because it requires an argument for the metaphysics as part of the support of the
epistemology.”
Douglas groups the eight sense of objectivity into three modes according to the type of processes they describe. They are:

Objectivity$_1$: Experimental Processes
1. Manipulable objectivity
2. Convergent objectivity

Objectivity$_2$: Individual Processes
3. Detached objectivity
4. Value-free objectivity
5. Value-neutral objectivity

Objectivity$_3$: Social Processes
6. Concordant objectivity
7. Procedural objectivity
8. Interactive objectivity

2.1. Experimental Objectivity$_1$

The first two senses of objectivity are grouped together in one mode because they relate to the interaction between the researcher and the natural world – the experimental process, broadly speaking.\textsuperscript{12} When scientists intervene in the world in predictable and repeatable ways or when they arrive at similar results from a variety of diverse attempts to capture the phenomena, they are more likely to arrive at objective descriptions of the natural world. These are, respectively, the manipulable and convergent forms of objectivity identified by Douglas.\textsuperscript{13} I believe that these and other methods for interacting successfully with the natural world are more often thought of as specifications of the requirement of empiricism rather than as senses of objectivity. The interaction between a researcher and the natural world occurs because of a commitment to empiricism. The specific processes (methodologies, practices, interventions) which allow for the reliable, repeatable and predictable interaction with the natural world are important as conditions for empirical success, rather than for objectivity. And there are surely more than two of these types of processes. It is worth noting that accounts of objectivity which fail to explicitly address

\textsuperscript{12} I have called this first mode experimental rather than “human-world interactive” (Douglas’ term) because I think it might otherwise be confused with interactive objectivity, which I discuss in great detail later in the chapter. The content of the mode remains the same.
experimental processes may still account for these features under commitments to empiricism. For these reasons, I will not spend much time discussing this mode of objectivity.

2.2. Individual Objectivity

The second mode of objectivity concerns the process by which individuals think and reason. Requirements that scientists distance themselves from their personal values are often conflated with the more stringent requirement that scientists excise all values from their reasoning. Douglas makes an important distinction between these two different requirements when she identifies detached objectivity and value-free objectivity. While detached objectivity requires a prohibition on using values in the place of evidence, value-free objectivity requires that all values are forbidden in scientific reasoning. Thus, the requirement of detachment amounts to an expectation that scientists not allow themselves to be led away from the results of empirical investigations by their own personal values and biases, while the requirement that individuals separate themselves from all individual values is much more demanding (perhaps impossibly so).

A third sense of objectivity thought to be related to the individual reasoning process is value-neutral objectivity, which requires that scientists make an effort to exclude “extremes” of values from their reasoning. In order to be value-neutral, scientists are expected to gain an awareness of the range of values that relate to their reasoning and then find a centered, balanced position. Each of these three senses of objectivity is distinct, and a normative philosophy of science will usually encode at least one of these individual processes for limiting values in science, depending on assumptions about the nature of reality and the relative importance of individual reasoning and social or community reasoning to the production of knowledge.

2.3. Social Objectivity

The third mode of objectivity concerns social or community processes. The outcomes of social processes are objective to the extent that they fulfill the requirements of one of the following three distinct senses of objectivity: procedural, concordant and interactive objectivity. I will discuss each of these senses in greater detail, since they form the basis for much of the discussion later in the chapter.
2.3.1. Concordant Objectivity

When we attribute objectivity to the results of scientific investigations because those results have been agreed upon by a variety of independent investigators, we have in mind *concordant* objectivity. In concordant objectivity, it is merely the agreement, and not the process that leads to agreement, that is relevant to the assessment of research. Polling researchers to see if they agree with the conclusions of a particular study would be one way of establishing concordant objectivity. Consensus conferences and traditional styles of creating guidelines in medicine would fall under this categorization. The strength of this sense of objectivity is that it allows scientific communities to bring the skills of a variety of scientists to bear on the results of research, and utilizes their individual judgements on such research. This distributes the effect of individual values on the results of research and allows the scientific community to assign greater objectivity to research that is widely accepted (for whatever reasons). The principal danger of this sense of objectivity is the possibility of group illusions – particularly those based on shared assumptions and theoretical frameworks. Wide agreement on research may indicate, for instance, that the range of theoretical frameworks being considered by scientists is very limited; that is, it may simply indicate a high degree of homogeneity in the scientific community. And while the strength of a particular theoretical framework may result in homogeneity, in practical terms narrowing the scope of research and restricting admission to the scientific community could achieve that same end. Silencing opposition and preventing dissenting views from being considered will be tempting routes to concordant objectivity. I will refer to this as the restrictive tendency. Insofar as a community of scientists is unable to assess whether the restrictive tendency has been in play, concordant objectivity cannot prevent group illusions from unduly influencing assessments. Moreover, any attempts to qualify concordant objectivity by adding requirements that group agreement be ‘rationally achieved’ or ‘under conditions of pluralism’ turns concordant objectivity into a version of interactive objectivity (discussed below).

2.3.2. Procedural Objectivity

---

14 Ibid., 462-3.
Procedural objectivity – sometimes referred to as mechanical objectivity – is attributed to the data produced by highly restrictive, and often quantified, rule-based procedures. In order to achieve objective results, we set up a procedure that leaves little or no room for input from particular individuals once it is implemented. Douglas gives the example of multiple-choice exams, which are procedurally objective because, given the rigidity of the rules governing the process of assessing grades on these exams, any individual can take the place of the exam grader and come up with the same result. This means that individuals using this procedure are interchangeable, and thus that idiosyncratic individual values will not normally impact the results of such a procedure. The strength of this sense of objectivity is that it uses a weakly social mechanism to eliminate individual values. Of course, this comes at a price, notably the restriction of the scope of knowledge to that which can be produced by such rigid procedures and, perhaps more insidiously, the imbedding of values into the design of the procedure in the first place.

Reliance on such procedures has historically garnered trust from the public. Just as numbers are often used to project certainty and confidence, procedurally objective results – often quantified – can be touted as ‘fact’ more easily than the rulings made on the basis of expertise. Theodore Porter suggests that the shift to procedural objectivity is often driven by a desire to secure the trust of others and to encourage the belief that subjectivity has been conquered, once and for all. Procedural objectivity relies on the implementation of a number of rules. “Rules are a check on subjectivity: they should make it impossible for personal biases or preferences to affect the outcome of an investigation.”15 Ultimately, since the individual perspectives are thought to have been expunged, “objectivity means the rule of law, not men.”16

2.3.3. Interactive Objectivity

Interactive objectivity is secured through discussion and debate within a community. The assumption behind this sense of objectivity is that the diverse perspectives of a variety of researchers can, under conditions of free and open debate, produce data that is as free from idiosyncratic individual values as possible. This relies on a Millian characterization of the

---

16 Ibid., 74.
benefits of open discussions (free speech), according to which we get at truth by subjecting ideas to the scrutiny of as many different perspectives as possible. This facilitates the discovery and correction of errors and falsehoods and greater understanding of the reasons behind well-accepted beliefs. Not only does the procedure of interactive discussion minimize the impact of individual values on research, if it is interpreted broadly it may also ensure that social values are minimized. This is because critical discussion proceeds on all aspects of research and is not limited to assessment of the results of research. Thus, the values and assumptions underlying the design of research from the earliest stages can be subjected to critique. If this is achieved, interactive objectivity appears to have a significant advantage over the two other senses of social objectivity, neither of which is able to acknowledge, much less address, the influence of social values on inquiry. Interactive objectivity is also distinct from procedural objectivity in its focus on community-level processes, as opposed to the processes or methods of particular research projects. Certain structural features of communities may be identified as key to creating the conditions conducive to such interactive discussion.

Like concordant objectivity, interactive objectivity can be compromised when members give in to the restrictive tendency and, influenced by various practical pressures, eliminate dissenting voices from their debates. It is often easier to discuss the merits of some piece of research when the individuals involved in the discussion all share a theoretical framework, background assumptions and commitment to shared cognitive goals and practical ends. Insofar as diverse perspectives make interactive debate and discussion more difficult, and agreement less likely, there will always be some motivation to place restrictions on the scope of the community and to minimize the scope of investigation. Thus, advocates of this sense of objectivity need to propose a mechanism for preserving diverse perspectives in order to ensure the critical debate is genuine. Without this mechanism, the restrictive tendency driven by a desire to make the debate proceed efficiently may undermine the benefits thought to accrue from interactive objectivity. If we are unable to tell whether a particular scientific community has fallen prey to the restrictive tendency, we will not know whether there is an unreflective, shared assumption or framework influencing research and as a result we will be unable to argue that interactive objectivity is managing (as best possible) the influence of social values on research. Thus, if the possibility for

---

17 The Millian framework is discussed in greater detail in Chapter Two.
18 Examples of such features were given in Chapter One.
the restrictive tendency to take hold is not addressed, interactive objectivity achieves no more than concordant or procedural objectivity. If it is, however, interactive objectivity has a great deal of appeal.

3. CCE Account of Objectivity

I turn now to an evaluation of objectivity in critical contextual empiricism.

3.1. Experimental Objectivity

CCE focuses on the need to control the influence of idiosyncratic preferences, biases, assumptions, interests and values, whether they are individual or social. CCE does not attribute any type of objectivity to the processes used by researchers to interact with the natural world. Rather, “[O]bjectivity is analyzed as a function of community practices rather than as an attitude of individual researchers towards their material or a relation between representation and represented.”¹⁹ This extends to methods used by researchers to obtain data about the natural world. Recall that CCE does contain a strong commitment to empirical adequacy (the “empiricism” part of CCE), the criteria for which overlap to a great degree with the requirements of manipulable and convergent objectivity suggested by Douglas. CCE is committed to empiricism, though Longino does not specify the details of the experimental interactions that will result in the most accurate empirical results, nor does she discuss empiricism as a condition of, or mode of, objectivity.

3.2. Individual Objectivity

According to the requirements of detached objectivity, scientists should ensure that their personal values do not usurp empirical evidence. CCE retains some commitment to this sense of objectivity. Again, though, it does so as part of its commitment to empiricism, not as part of the account of objectivity. And I think this is right: it is a commitment of empiricism that scientists not override empirical evidence with values. To call this commitment to empiricism detachment

is misleading, since it suggests that scientists are capable of splitting themselves into different selves, the scientist version of which is detached from person values (the slide to value-free objectivity here is a quick and easy one). Further, it suggests that the practice of science can be separated from any and all values through the will-power of individuals. Longino explicitly rejects these additional commitments when she defends CCE, but it is worth noting that her commitment to empiricism covers this sense of objectivity as it has been defined by Douglas. As one might expect, given that her account is called critical contextual empiricism, she is not interested in having scientists reject empirically adequate data because of values. But she is also not interested in taking any further step toward value-free objectivity.

In fact, Longino explicitly rejects value-free objectivity:

The objectivity of individuals in this scheme consists in their participation in the collective give-and-take of critical discussions and not in some special relation (of detachment, hardheadedness) they may bear to their observations.\(^{20}\)

Though Longino does not get into this, there is an additional concern about what exactly a value-free individual might look like, and whether the concept makes any sense when applied to real human beings. Consider, for instance, a medical researcher who is making a decision about which disease to investigate in her research. She is not value-free if she chooses to investigate one disease rather than another because that disease affected, or even killed, a friend or loved-one. She is not value-free if she chooses to investigate a disease affecting impoverished people in developing nations on the basis of concerns about social justice and maximum benefit. And, with respect to social values, she is not value-free when she chooses which disease to research based on the available funding for research as set out by the national research institute in her home country. So what would value-free decisions at this stage look like? The difficulty in characterizing a process of value-free reasoning persists in cases where we examine choices of research method, of outcomes, of which journal to publish in (or whether to publish) and so on. That is, it persists at every stage from the very first to the very last choice made in the research

\(^{20}\) Longino, *Science as Social Knowledge*, 79. In a similar statement in another paper she writes, “Knowledge and objectivity, on this view, are identified as the outcomes of social interactions and, hence, are located not in individuals but in communities. Individuals must participate in these interactions in order that knowledge be produced; but their objectivity consists in such participation and not in any special cognitive attitude (for example, impartiality or distance) they bear to proposed objects of knowledge.” Longino, “Essential Tensions – Phase Two,” 268.
process. It is no less a value-laden choice when the researcher aims to do the right thing and choose the most judicious topic, method, outcome, or journal, than when she aims to act self-interestedly and chooses the topic, method, outcome or journal based on personal gain. Some of these choices are laudable and others are not, and we have good reasons for making these assessments – but they are all underpinned by values. It is not clear that we can make any sense out of the claim that scientists can learn to reason in a wholly value-free manner. Values pervade all aspects of knowledge production. We had best get on with dealing with those values.

In a statement consistent with the comments I have just raised, Longino rejects the types of individual objectivity that attempt to characterize the ways in which an individual’s reasoning processes might be more or less objective:

Knowledge, in the view I will advocate is produced not by individuals as individuals but by communities – that is, by individuals in interaction with each other. This means that normativity, if it’s possible at all, must be imposed on social processes and interactions, that the rules or norms of justification that distinguish knowledge (or justified hypothesis acceptance) from opinion must operate at the level of social as opposed to individual cognitive processes.\(^1\)

It is not fruitful to talk about the objectivity of individuals. This position is congruent with the commitments of feminist epistemologists more generally, who tend to stress the inescapable subjectivity of all human knowers and hence the inappropriate nature of attempts to mask or suppress our subjective perspectives. The concern with an overemphasis on individual objectivity is twofold. First, it is not usually, or perhaps ever, achieved. Extensive feminist philosophical investigations lend support to this concern, as do the conceptual problems I have just raised. Second, even if any of the senses of individual objectivity were to be achieved, this would not rule out the influence of social values on the investigation and so may at best be considered necessary – though not sufficient – for objectivity. And yet the trend in the literature aiming to ‘fix’ value-laden science is often in favour of greater emphasis on detachment and mechanisms designed to increase the value-freedom of individual perspectives. This is dangerous when it obscures the influence of social values on science. I will discuss this danger further in the next section of the chapter. Concerns about the persistent role of social values in scientific research are summed up by Longino as follows:

The transformation of an idea into scientific knowledge has the effect of purging it of idiosyncratic features of its initial proponents. This gives it an impersonalism often misinterpreted as objectivity. It is not the impersonality, however, but the collaborative social process of transformative interrogation that makes it objective. And while the marks of the individual may be eliminated by this process, the marks of culture are not.\(^2\)

According to these descriptions, Longino’s account of objectivity is focused entirely on *social objectivity*; she argues that social mechanisms can be used to limit the influence of values on knowledge. Her social account of objectivity is meant to replace, rather than complement, an individualist account.\(^3\)

### 3.3. Social Objectivity\(^3\)

CCE is based on a social account of objectivity. Of the three possible social accounts, the first two are rejected in favour of the third. I will outline the reasons for this choice below.

#### 3.3.1. Concordant Objectivity

Concordant objectivity is afflicted with a familiar problem. Recall that concordant objectivity is achieved when there is agreement within a community. Because this form of objectivity stops short of specifying further criteria for the process leading to agreement, it is vulnerable to criticisms about potential abuses. If all that matters is agreement, and not how that agreement was achieved, there is a significant potential for distortions based on power, authority and influence – as there would be in any human community under this relatively weak requirement. Powerful community members may force or manipulate others into agreement under this account. Specification of the procedure leading to agreement seems more likely to bring powerful forces to light, provided the procedure includes conditions designed to ferret out and critically examine such forces. Unfortunately, as noted earlier, attempts to save concordant objectivity by specifying that the process leading to agreement should be rational or pluralistic turns concordant objectivity into interactive objectivity. For these reasons, Longino is not an

---


\(^3\) Longino, “Subjects, Power and Knowledge,” 113.
advocate of concordant objectivity. This likely is not a big surprise, given that feminists are generally concerned about the sorts of oppression that can result from precisely these sorts of institutional arrangements. I turn now to procedural objectivity, which is much less easily defeated and also much more common.

3.3.2. Procedural Objectivity

Procedural objectivity is appealing due in no small part to the democratic principles it appears to encode. If we can set up an objective procedure, then we can rely on the results of this procedure without having to waste time analyzing the outcomes. And the people who actually carry out the procedure are fully interchangeable and are prevented from contaminating the results with their idiosyncratic values. There are costs to this, however. In order to set up a procedure we need to make a number of assumptions about the nature of the knowledge we intend on getting, and the types of questions and methodologies that will get us there. This is the problem of hidden background assumptions. Drawing on the example of the multiple choice exam offered by Douglas, while such exams can be reliably marked by any number of evaluators, they can only produce results restricted to the set of questions that can be fit into the multiple-choice format. And, as all teachers and professors know, a student may appear to be quite knowledgeable on a multiple exam and significantly less so on, for instance, an essay exam. If we think of the variety of formats available for testing an individual’s skills, the multiple choice exam is useful in some contexts, for some questions, and any attempt to suggest that it should be used to evaluate all skills in all disciplines would be ridiculous (imagine someone testing bedside manner or surgical skill in this way) no matter how “objective” it might be. Multiple choice exams and other strict procedures for evaluating information rely for their proper application on an appreciation of the assumptions upon which they were constructed.

Because social values may be encoded in, among other things, the choice of question, intervention or outcome, the threshold for significance, and more broadly in the expectation that complex, difficult or obscure factors will be (and should be) eliminated within a particular procedure, multiple choice exams and other procedures thought to produce objective results are often inaccurately portrayed as socially objective.24 No matter how successful such

methodologies and rules are in preventing individual values from influencing the collection or analysis of observations, there remain countless more insidious ways in which contextual values influence scientific research. The false sense of objectivity created by the use of strict rule-based procedures may do harm when it leads to claims to objectivity in cases where social values persist. This false sense of objectivity has been of particular concern to feminist scholars as it has often led to an inappropriate trust in science and scientists, to the detriment especially of those members of oppressed groups who have traditionally had the least authority to challenge the results of such science. Longino is interested in developing an account of scientific inquiry that acknowledges and deals with social values, and which permits for the critical evaluation of hidden background assumptions. Procedural objectivity alone will not suffice for this project.

### 3.3.3. Interactive Objectivity

With respect to the modes of objectivity identified by Douglas, CCE focuses exclusively on social objectivity – and in particular on interactive objectivity. Longino states that it is only social processes that matter in determining the objectivity of scientific research. Social objectivity is not adopted in conjunction with commitments to experimental and individual objectivity but in opposition to these other senses. So the interesting and potentially innovative move by Longino, then, is the rejection of the need for discussions of individual objectivity and the replacement of individual objectivity with a socialized, contextualized, interactive objectivity. Perhaps more to the point for this analysis, Longino is very critical of procedural objectivity and outlines the ways in which such an approach falls short. Longino’s biggest concern is with the dangers of unexamined background assumptions in scientific inquiry, since they often encode social values. Procedural objectivity can exacerbate this problem, since it masks social values under the guise of objective procedures and rules, as we have just seen. This does not mean that there is no place for procedures designed to secure objectivity, only that the limitations of such procedures must be appreciated.

The products of scientific communities are objective to the extent that the community upholds four norms: avenues for criticism, responsiveness to criticism, shared public standards and diverse perspectives. The four norms create the conditions for critical community

---

25 This is discussed further in Chapter Two.
discussions. The transparent disclosure of a diverse set of subjective values combined with certain criteria for open, critical, responsive debate contribute to the interactive objectivity of knowledge claims made by a community. These conditions make it more likely that the background assumptions underlying knowledge claims in the community have been subjected to transformative criticism. Longino writes,

[T]he greater the number of different points of view included in a given community, the more likely it is that its scientific practice will be objective, that is, that it will result in descriptions and explanations of natural processes that are…less characterized by idiosyncratic subjective preferences of community members than would otherwise be the case.

There will be context-specific requirements for this sort of objectivity; it will be necessary to specify the appropriate avenues for debate, the sorts of diverse perspectives that will be attended to, and the appropriate weighting of different epistemic values. Determining the particular structural arrangements that will best meet these requirements is a matter for empirical investigation. Setting out the requirements in the first place is a philosophical task, and this is what CCE does.

4. Concerns

In response to this defence of interactive objectivity, a critic might point out that procedural and concordant objectivity are not necessarily incompatible with interactive objectivity. We can respect other forms of social objectivity without compromising our commitment to interactive objectivity or social epistemology more broadly. Procedures (procedural objectivity) alone are problematic, but within a larger context of interactive objectivity, they might have a role to play. In the original article in which the eight senses of objectivity are elucidated, Douglas argues that each sense is unique, and that we should embrace the irreducible complexity of objectivity. Isn’t Longino subverting this inclusive project by suggesting that, not only are all senses of individual objectivity problematic, but also that most forms of social objectivity – with the exception of interactive objectivity – fail as well?

26 Longino refers to this as contextual objectivity.
Strict rule-based procedures create the conditions for trust. People are more inclined to understand and accept the results of such procedures. While advocates of procedural objectivity describe this as factoring out the particular researcher (by making researchers interchangeable), another way of describing the situation is to say that we are enhancing the conditions of interactive objectivity by making our research as publicly accessible as possible. The results of rule-based procedures are not yet objective until they have been scrutinized by a variety of researchers from diverse backgrounds, but the character of the procedure may make it more likely that diverse researchers can engage in the discussion. So we no longer decree the results of these procedures objective in virtue of the way they were produced, but we can acknowledge that some procedures make it easier for other researchers to engage with and critically evaluate research. So it can be useful to set up rule-based procedures insofar as they make the interactive evaluation of results more likely to occur. In this way, procedural mechanisms are subsumed within an account such that offered by CCE. To reiterate, however, we reserve the honorific title of “objective” for the results of interactive debate.

Marianne Janack has suggested that attempts to isolate the core meaning of objectivity will always miss out on some important aspects of this ‘variegated motley’ ideal. Thus, she rejects Robert Nozick’s attempt to equate objectivity with invariance, Arthur Fine’s attempt to isolate trust as the central goal of objectivity, and Longino’s attempt to suggest that objectivity is the label we assign to the products of transformative community debate. According to Janack, these attempts are problematic because they disguise the fact that “a vast and disparate array of virtues, goods, and procedures are collected under the umbrella of the concept.”28 The interesting and potentially fruitful project, she argues, would be to take apart the concept and highlight the tension (and perhaps even inconsistencies) between different senses of the concept. So while Longino’s account is useful in that it contributes to our overall understanding of objectivity, Janack suggest that the account is too concerned with unification, or, at the very least, isolating the core important element of objectivity.

To add to this, one might question whether paring down the concept of objectivity in this way puts extra pressure on Longino’s criterion of empirical adequacy. Is this concept going to have to ‘pick up the slack’, as it were, for the slimmed down account of objectivity? We have already seen the ways in which Longino leaves elements of experimental and individual

---

objectivity out of her account, preferring to leave discussion of such things to the debate over empiricism. How much can we build into empiricism and is this division of labour between justification and empirical adequacy problematic in any way? I think the burden of proof is on Janack and others to show that Longino’s empiricist criterion cannot contain everything assigned to it. I have not seen a full argument to that effect yet, so at this point I am content to say that Longino’s characterization of objectivity is a good one. However I will certainly be attentive to any arguments made on this matter in the future, since I think it could potentially become a problem.

5. Objectivity and Values in Medical Research

What is the overriding concern in science today; what, in Daston and Galison’s terms, is the most dangerous form of subjectivity in modern scientific inquiry? According to their historical analysis, scientists in the late twentieth century are preoccupied with the corrupting influence of individual idiosyncratic values. Scientists have recognized that they have to take steps to ensure that individual values do not limit their ability to understand the natural world. Because of this, there is a lot of attention to individual objectivity: we are all familiar with the requirement that scientists aim to be value-free (in Douglas’ terms). In addition, scientists have developed strict methodologies and procedures in order to prevent any remaining individual values from having a distortional effect on the results of scientific research. The trend in science is to rely upon scientists to be value-free and methodologies to be procedurally objective. This two-pronged approach to the objectivity of scientific inquiry is noticeable in the medical context. I will be paying special attention to the reliance on procedural objectivity. Before turning to this investigation, it is worth taking a moment to discuss the possible reasons for this shift to procedural objectivity.

5.1. Why Procedural Objectivity?

In his book *Trust in Numbers: The pursuit of objectivity in science and public life*, historian Theodore Porter outlines various political and sociological factors that led to the current reliance
on procedural objectivity in certain areas of science.\textsuperscript{29} In particular, Porter focuses on two factors that have been particularly influential: lack of trust and contested boundaries. With respect to the former, he draws attention to the lack of public trust in the personal judgements of scientists, and the desire of scientists to overcome this distrust. According to Porter, methods and rules thought to be objective often serve as an alternative to trust in the sciences.\textsuperscript{30} Alfred Tauber has written persuasively about the erosion of trust in medicine over the past half century, and the considerable “crisis in trust” facing medicine today.\textsuperscript{31} Drawing on Gallup polls, he offers evidence that trust in medical professionals – particularly physicians – has been slowly eroding over the past fifty years in the USA. Similar trends have been noted in Canada.\textsuperscript{32}

With respect to the latter, Porter reveals that scientists working in contested domains, such as medicine, have often attempted to gain (or regain) the respect of other scientists, and the public, by relying heavily on inflexible quantitative methods – so-called objective procedures.\textsuperscript{33} Heavy reliance on procedural objectivity is most common in disciplines that lack clear boundaries, and which are engaged in defending insecure boundaries against outsiders.\textsuperscript{34} Even a brief history of medicine in North America indicates a persistent negotiation (often a rather cutthroat ‘negotiation’) over the boundaries of the medical domain. Today, physicians defend the boundaries of the medical domain from “web doctors”, government regulators, administrators, and alternative medical practitioners. In these sorts of contested disciplines, “We find there is a pervasive dread of ‘the prejudice of the investigator’, often a willingness to leave untouched the most important issues in order to deal objectively with those can be adequately quantified.”\textsuperscript{35} In medicine, we see reliance on the results of research thought to be most objective, even when this means neglecting fruitful evidence lower on the evidence hierarchy. With this brief historical overview in mind, let us consider the impact of EBM on methods of securing objective evidence.

\textsuperscript{29} Porter uses the term mechanical objectivity for what I (following Douglas) call procedural objectivity.
\textsuperscript{30} Porter, Trust in Numbers, 228.
\textsuperscript{31} Tauber, “In Search of Medicine’s Moral Glue,” 42. Putnam’s influential book Bowling Alone: The collapse and revival of American community is another good source on the erosion of trust in American society generally.
\textsuperscript{32} Kondro, “Threats to Medical Professionalism Tackled in Canada,” 316.
\textsuperscript{33} Porter, Trust in Numbers, 211.
\textsuperscript{34} Contrary to what many people might suspect, relatively secure disciplines often eschew strict reliance on procedures and methods designed to promote objectivity, and are more open-minded about the importance of creativity, intuition and expert judgement in their domain. Porter, Trust in Numbers, 230.
\textsuperscript{35} Porter, Trust in Numbers, 229.
5.2. The Procedural Objectivity of RCTs

Sociologists Stefan Timmermans and Marc Berg have argued that the EBM movement was driven largely by a desire for standardization in medicine. In the late 1980s a focus on standardization (re)emerged in healthcare. In their description of the changes that came about in this period, Timmermans and Berg write, “of all the kinds of standardization attempts that affected medicine in the twentieth century, evidence-based guidelines represent the farthest reaching and most direct attempts to prescribe and preset the actions of healthcare professionals.”36 In contrast with longstanding attempts to institute terminological standards (such as international classifications of diseases) and performance standards (such as a maximum level of complication rate for an operation), the EBM movement proposed procedural standards (including clinical practice rules and guidelines for physicians, and the evidence hierarchy for researchers).37 This form of standardization relies on a procedural account of objectivity.

Top ranked research methods, such as the RCT, are expected to produce maximally objective results. It is not difficult to detect the underlying commitment to procedural objectivity in the evidence hierarchy. You can replace the team of researchers with any other team and get the same results if they were to perform the same RCT. So research methods, particularly RCTs, are thought to be best at getting objective results. And the idea is that as you work your way down the hierarchy, the research methodologies are less and less able to rule out the idiosyncratic values of the researcher. The best procedures are those that eliminate the individual researcher to the greatest extent possible. It comes as no surprise, then, that qualitative methodologies are not highly valued. A reliance on procedural objectivity in medicine indicates great concern with certain specific individual values that may influence research during the course of trials, but very little attention to social values or individual values that may have an influence before and after the trial.

5.3. The Evaluation-based Education Analogy

37 Ibid., 25.
Consider the following analogy. A group of teachers and administrators from a highly-regarded teaching institution decide that there is a problem with the state of education in Ontario. Let’s call them the Evaluation-based Education (EBE) Working Group. They develop a new way of evaluating the effectiveness of public education. They propose a hierarchy of exam types which elevates multiple-choice exams as the gold standard, places short answer questions in the middle of the hierarchy, and ranks essay exams, oral exams and in-class and informal discussions at the bottom as means of evaluating students. Whenever possible, multiple-choice exams are recommended, and it is even better if they can be standardized and used throughout the province. Teachers are advised to use the top-ranked methods for evaluating their students, and then results are compared across classrooms and schools. While the studies yield information about how individual students are faring, and how certain sub-groups within schools and classrooms are doing, this is not the information of most interest to the EBE group. Rather, attention is focused on the average student data, which is used to determine whether the education received by students overall has been effective. This hierarchy is justified in part on the basis of arguments about the relative objectivity of different evaluation techniques. Multiple-choice exams can be reliably marked by any evaluator, and a machine can even do it if we want to remove the evaluator completely. Thus the influence of individual values (of the marker/evaluator in this case) can be as diminished as possible. Or so it seems.

In a climate of pervasive evaluation-based education I suspect we would find educators asking themselves two questions. First, are we necessarily getting at more objective results when we use the best methodologies according to the hierarchy devised by the EBE group? And second, what are we trading off in order to gain objectivity? This first question is vital because the multiple-choice exam only appears objective because the evaluator has little control over the particular outcome during the period of evaluation. But the evaluator has a great deal of control

---

38 This sort of movement is already happening in education. The What Works Clearinghouse (WWC) formed by the Institute of Educational Sciences in the United States, for instance, provides research evidence to support their recommendations on educational reform. The evidence is presented in a surprisingly medicalized manner (though it would be judged to be of very low quality by EBM standards). The WWC standards of evidence have two criteria for quality of study design: “RCT (meets evidence standards) or QED (meets evidence standards with reservations).” Also considered in the overall evaluation are: effect size, sample size and statistical significance. It is not possible to assign an intervention the top rating of “positive effects” without having evidence from at least one RCT. http://ies.ed.gov/ncee/wwc/reports/beginning%5Freading/success%5Ffor%5Fall/effectiveness.asp. Similarly, the “No Child Left Behind” program in the United States is built on four principles: “accountability for results, more choices for parents, greater local control and flexibility, and an emphasis on doing what works based on scientific research.” Emphasis added. http://www.ed.gov/nclb/overview/intro/edpicks.jhtml?src=In.
over the particular outcome when they decide which material should be tested in the first place, choose to put certain questions and not others on the exam, curve the grades after the fact, or decide to exclude students who received 0’s from the final calculation so as not to distort the overall results. It does not look like we are achieving objectivity in any robust sense merely by applying the hierarchy. Moving to the second question, in the educational context it seems clear that we are trading off our ability to evaluate certain important skills that students have (such as writing skills) in order to have more objective evaluations of other skills (such as comprehension). If these standards of evaluation are persistently applied it follows that we would have very little access to information about how well our students are learning to write (assuming we still bother to teach it at all). And yet we might think that it is important to evaluate a wide range of skills, and that we do a disservice to our assessment of students when we limit our evaluations in this way.

The answer to the first question above – are we necessarily getting at more objective results when we use the best methodologies according to the hierarchy? – is no. I presented an argument to this effect in Chapter Five. That makes the second question even more important, since we are trading off all sorts of things and not even getting what we think out of the deal. We do not necessarily get more objective results, so what are we paying for this privilege? I will turn now to the question of trade-offs.

5.4. Economic Models and RCTs

Nancy Cartwright has recently drawn an analogy between RCTs and models in economics.\(^3\) Cartwright argues that models in economics are “front-loaded”. All sorts of substantive (and value-based) assumptions are made in the design of economic models, and if those assumptions are correct the models deliver remarkably consistent – even, she claims, deductively valid – results.\(^4\) That is, they are extremely good at establishing the internal validity of their results. There are costs to this however. We pay the price with two distinct types of trade-offs. The most obvious is the trade-off between internal validity (or rigour) and external validity (or real world

\(^3\) Cartwright, “Gold is where you know what you need to know.”

\(^4\) Cartwright, “The Vanity of Rigour in Economics: Theoretical Models and Galilean Experiments.” Cartwright is mistaken in believing that these models can ever be deductively valid. At best, such models seem to establish their conclusions through eliminative induction.
applicability). We can call this the trade-off for rigour. If a model or method must be highly idealized or rely on a number of critical assumptions in order to produce deductively valid outcomes, it is often inappropriate to assume that results and outcomes can be applied beyond the scope of the initial conditions. Moreover, if it turns out that any of the assumptions propping up such a model or method is suspect, we are left with highly contested results with little applicability.

Returning to the similarities between economics and medicine, then, the results of RCTs, as with economic models, can be very strong but they come at a high input cost. If any of the assumptions turns out to be shaky, the validity and applicability of the results is compromised. In chapters four and five I argued that we have good reason to question assumptions about the epistemic superiority of RCTs. The second trade-off we make is that we give up investigating other types of questions. This is an intellectual trade-off. We could investigate the economic situation in a way that would give us personal narratives as data, but we choose to investigate it in a way that gives us numbers. Similarly, we give up investigating biological mechanisms of disease or sources of physician-patient miscommunications for the sorts of questions that can be asked in a RCT. This diminishes the range of research we perform and the range of treatments we offer. We will always have to make trade-offs but if we do not explicitly recognize that this is what is happening, we cannot subject the decisions to critical analysis. Transparency is the first step to accountability when making these sorts of decisions. It is precisely transparency on trade-offs that is not discussed within EBM.

5.5. Method and Methods

The danger of relying on procedural objectivity, or objectivity of method, is that a (legitimate) focus on practice or process (method, broadly construed) shifts to a (somewhat narrower) focus on particular research trial designs (methods). Thus, scientists come to believe that a particular research method or trial design fully secures objectivity, and thus end up neglecting the other elements of scientific practice and process – or method, broadly construed – (such as critical debate) that are also significant in contributing to objectivity. This highlights the distinction between procedural objectivity in medicine, which focuses on particular research methods, and interactive objectivity, which focuses on community-level structural features. What is needed in
medicine is greater attention to social structures of knowledge-production. The conditions of interactive objectivity could enhance the production of knowledge in medicine.

6. Interactive Objectivity and Medicine

The version of critical contextual empiricism I offered in Chapter Two proposes four norms designed to ensure adequate critical debate within knowledge-productive communities. The third and fourth norms – *shared public standards* and *cultivation of diverse perspectives* – serve to define the boundaries of the community. The requirement that communities recognize the perspectives offered by qualified others serves to protect diversity, while the requirement of shared standards protects unity. These two requirements push against each other and the tension between the two shapes the boundaries of particular communities. These norms play an important role in ensuring the interactive objectivity of knowledge produced in a community. This role was discussed in detail in chapters two and six. In addition, the first requirement – *recognized avenues for criticism* – provides direction on how to ensure interactive objectivity (and *responsiveness to criticism*) through social mechanisms. In what follows I will discuss specific recommendations for the medical community suggested by the first CCE norm.

6.1. Recognized Avenues for Criticism: The Open Science Movement

According to CCE, there must be *recognized avenues for criticism* within knowledge-productive communities. In scientific communities, this typically includes peer-reviewed journals, conferences, workshops, and so on. What I will call the “open science” movement of recent years provides a number of specific recommendations for ensuring the transparency and publicity of scientific research. In other words, it provides recommendations for ensuring recognized avenues for criticism and for ensuring community members make use of these avenues. These include: rejection of non-disclosure clauses in research contracts, disclosure of competing interests in publications, prohibitions on ghost authorships, a mandatory clinical trials registry, open peer-review, open access journals, open submission processes, public funding for research, an emphasis on original research, greater awareness of the collective nature of medical
decision-making, and more comprehensive analytic training for health professionals. I will
discuss each of these recommendations with respect to the medical context.

6.1.1. Regulation of Research Contracts

Medical researchers often receive funding for their trials from pharmaceutical companies.
Because these companies have an interest in maintaining control over the data produced by the
trials they sponsor, they sometimes require that researchers sign contracts in which they agree
not to discuss research findings unless approved by the company. These are referred to as non-
disclosure clauses in research contracts. For example, in the mid 1990s, Nancy Olivieri, a
researcher at the University of Toronto and The Hospital for Sick Children, undertook a trial
funded by Apotex Inc. (a pharmaceutical company). As part of her contract, Olivieri agreed to
the following non-disclosure clause:

All information whether or not obtained or generated by the investigators during the term
of this agreement and for a period of one year thereafter, shall be and remain secret and
confidential and shall not be disclosed in any manner to any third party, except to an
appropriate regulatory agency for the purposes of obtaining regulatory approval to
manufacture, use or sell [the drug] unless the information has been previously disclosed
to the public with the consent of Apotex. The investigator shall not submit any
information for publication without the prior written approval of Apotex.  

Reference to “any third party” here includes patients. At the time Olivieri signed the contract,
such non-disclosure clauses were common. The complexities of the Olivieri case have now filled
several hundred pages in many books and publications, and it is not my intention to repeat those
descriptions or analyses here. What I would like to do is draw attention the problematic nature of
non-disclosure clauses. A cursory overview of the case will be helpful to this end.

Medical professor and researcher Nancy Olivieri entered into a contract with Apotex
agreeing to the clause just cited (among others). Part way into a trial on a new drug she grew
concerned about a significant number of patients who appeared to be doing worse under the new
treatment. Representatives from Apotex disagreed with Olivieri’s interpretation of the
preliminary data and advised her not to convey her concerns to the patients in the trial. When
Olivieri did so against their wishes they terminated the trial and pursued legal action against her.

41 Quoted in Shafer.
The case was resolved in favour of Olivieri many years later. The Olivieri case drew attention to the decrease in control over research data by primary researchers and the potential for significant harm when researchers are not allowed to convey concerns to their patients or talk about these concerns with colleagues. What non-disclosure clauses do is shut down avenues for critical discussion. The Olivieri case highlights the epistemic damage done by such agreements and the potential harms that arise. As of March, 2001, the teaching hospitals affiliated with the University of Toronto have harmonized their policies on research contracts so that contracts are more carefully analyzed for potential conflicts of interest. In light of the scandal surrounding this internationally recognized case, researchers have become more aware of the dangers of signing confidentiality agreements with pharmaceutical companies. Non-disclosure clauses are widely regarded as damaging to the scientific process.\footnote{I occasionally speak about the dangers of “corporate interests” in science. To clarify, I am concerned with those cases in which corporate interests impinge on the processes integral to good scientific practice. I believe that to be the case with non-disclosure clauses.}

\subsection*{6.1.2. Declaration of Competing Interests}

In a relatively recent development, all of the top medical journals now require researchers to declare any competing interests or potential conflicts of interest they have with respect to the research they are submitting for publication. So, for example, if a trial is funded in part or entirely by those who stand to profit from positive results, this must be disclosed upon publication. Requirements of disclosure are helpful at highlighting potential biases, though it is unclear exactly what action the reader of such a publication is meant to take, aside from taking a somewhat more critical attitude toward the results.

\subsection*{6.1.3. Prohibition on Ghost Authorships}

Ghost authoring occurs when the names of researchers who did not contribute substantially to a trial are included in the list of authors at publication. Ghostwriters are professional writers who are paid (usually by pharmaceutical companies) to write up the results of medical trials. These ghostwriters do not retain any credit for their work in the final publication, but pass that credit along to others (ghost authors) for a price. Often pharmaceutical companies will look for well
known and highly regarded researchers ("Key Opinion Leaders" or "Thought Leaders") whose names they can attach to a particular publication as ghost authors. Ghost authoring is reportedly widespread in medicine. The "thought leaders" who agree to attach their name to a publication are often restricted from viewing the original data from the trial. Instead they are given only already-analyzed and carefully selected data: that which reflects most positively on the treatment. What this means is that researchers acting as ghost authors are unable to offer any critical perspective on the data they are claiming to have produced. This significantly limits the opportunity for genuine critical discussion on the results of research. Medical journals have taken steps to limit this practice by requiring that authors on publications explicitly describe their level of involvement in the research, but the practice is still common.

6.1.4. Clinical Trials Registry

In order to ensure recognized avenues for criticism, the medical community has to extend critical social mechanisms beyond requirements placed on researchers, reviewers and journal editors. One of the concerns raised in recent years has been with the selective reporting of the results of trials. Articles finding a positive effect for a new treatment are more likely to be published than those with negative results. This is heightened in the case of positive studies for pharmaceutical drugs where the pharmaceutical company sponsored the research. In response to this trend toward suppressing negative results, critics have argued for the importance of public registration of research trials. Over the past few decades, the support for a public clinical trials registry has grown, though most versions today remain voluntary and piecemeal and so have a limited influence. In 2000, the National Research Register was established in the U.K. In 2004, the House of Delegates of the American Medical Association called for the comprehensive registration of all clinical trials conducted in the United States through the National Library of

---

43 Laine and Mulrow.
44 Ibid.
45 Bekelman et al.; Sackett and Oxman.
46 The voluntary registry set up by the Association of the British Pharmaceutical Industry (ABPI), for instance, has been only marginally successful in posting information about ongoing trials in the U.K.: "The completeness of the information varies widely and often comes nowhere near what was promised in the initial ABPI press release." Similarly, www.clinicaltrials.gov, the website set up by the American National Institutes of Health, has been only marginally successful in getting researchers to register their trials. See Rennie, "Trial Registration" for details.
Medicine website.\textsuperscript{47} Registration was to be monitored by the ethics boards charged with approving trials. As of 2004, the International Committee of Medical Journal Editors states that, “trial registration is largely voluntary, registry data set and public access to them vary, and registries contain only a small proportion of trials.”\textsuperscript{48} This is in spite of what these editors believe to be the significant advantage of trial registration for the medical community.

Pharmaceutical companies including GlaxoSmithKline (GSK), Merck & Co., Johnson & Johnson, and Eli Lilly & Co. have all indicated some willingness to post ongoing studies online. While there is some resistance, the pharmaceutical industry is struggling with its public image in light of various recent drug scandals and is more open to the suggestion that might otherwise be expected. Of course, there are a variety of proposed loopholes to registration including registering only late-stage clinical trials or trials after approval or publication. The deputy editor of \textit{JAMA} firmly states that exceptions and qualifications of these sorts are inappropriate. There is no use for a registry unless it is: adequately funded, mandatory, designed to include all trials, adequately policed, equipped with stiff penalties, and most of all independent.\textsuperscript{49} A registry that meets these criteria would go some way to providing the information required for a proper critical evaluation of evidence. It would help prevent unnecessary duplication of research and fulfill ethical commitments to research participants who should have access to the results of research in which they participated. Such a registry would have to be supported by Research Ethics Boards, journal editors and individual researchers in order for it to be successful. The World Health Organization has set up the WHO International Clinical Trials Registry Platform which may serve this purpose in the future. Currently the standard appears to be the American website: www.clinicaltrials.gov.

\textbf{6.1.5. Open Peer-review}

Of the avenues for critical debate within medicine, peer-reviewed journals are primary. Peer-review of submissions to medical journals allows for a range of outside perspectives on the methods and results of research, the value of the questions asked and the quality and relevance of the findings. However, peer review can also be a mechanism for exclusion, particularly when

\textsuperscript{47} \url{www.clinicaltrials.gov}
\textsuperscript{48} De Angelis et al., 1363.
\textsuperscript{49} Rennie.
blind evaluation is circumvented. Blind evaluation takes place when the identity of authors is masked so that reviewers do not know whose research they are evaluating. Blind evaluation was originally thought to prevent conflicts of interest during review. The concern was that reviewers might be influenced by their relationship, whether collegial or competitive, with the authors of the papers they are reviewing. Recently, there have been a number of scholars who have argued that sufficiently motivated reviewers have no trouble bypassing mechanisms of blinding. Because blind evaluation is not actually occurring much of the time, it is not a particularly efficient mechanism. Transparency, or unmasked review, has been a popular topic of discussion in the past few years. Transparent, or open, peer-review can aid in the appropriate evaluation of research since it allows reviewers to contextualize the findings (“Oh they’re from so-and-so’s lab/clinic, I see why they have done it this way”) and offer more specific and helpful comments. It also prevents reviewers from hiding behind their anonymity and encourages deeper critical engagement among researchers because comments and responses can go back and forth for some time. Of course, this is not to say that it would be unproblematic. Certainly there will be researchers who take the opportunity to unfairly critique the work of new researchers or old rivals. However, 1) there are reasons to think that this is happening already, and 2) there are additional mechanisms in the open science movement that may help address these problems.

The recent open peer-review experiment undertaken by Nature in 2006, for instance, was not only transparent about authors and reviewers, it allowed any other members of the research community to add comments to the submitted manuscript online, as long as they identified themselves when doing so. Traditional peer-review was complemented by online comments, and editors used both to make final decisions. While researchers were, on the whole, not that interested in open peer-review (only 5% of authors opted to post their manuscripts for public review) the relative unpopularity of the open process is subject to change and does not bear on the epistemic value of such a process. In fact, all of the authors who received additional online comments found them to be at least somewhat valuable. The main concern seemed to be with lack of comments, since many of the papers did not receive additional comments online. The editor of Nature suggested that researchers were too busy and “lacked sufficient career

---

50 Occasionally this is reversed so that the reviewer is known to the authors but the authors remain unknown to the reviewer. This is still considered (single) blind evaluation.
51 Falagas, Zougkas and Kavvadia; Yankauer.
52 Editorial, “Peer Review and Fraud: Two assessments of the refereeing process highlight challenges for journals.”
incentive” to spend time commenting on the work of peers online.\textsuperscript{53} Further, he suggests that attitudes may be different if researchers are asked to comment on already-published papers and he proposes to investigate this possibility. If researchers were more interested in critical engagement after publication, this would still be a great contribution to the interactive objectivity of research since publication is no strict marker of knowledge. Further empirical investigations on the most effective form of peer-review are warranted.

6.1.6. Open Access Journals

A number of medical journals have instituted policies that would improve the availability and transparency of research to members of the medical community. Most medical journals have sections dedicated to opinion pieces and editorials, and physicians often take advantage of these venues in order to pursue arguments with colleagues. In addition, many journals now offer a “rapid response” feature, which allows researchers to provide immediate electronic feedback on new research, with minimum effort. Taking this a step further, open access journals provide online content free of charge to anyone regardless of institutional affiliation. Such journals not only provide an location for critical discussions, they ensure that all members of the medical community have access to these discussions, as do the members of other communities (including CAM communities) and the public. This contributes to the diversity of perspectives involved in medical debates and increases avenues for critical discussion.

6.1.7. Open Submission Process

Medical researchers often submit their research to top medical journals, only to be rejected. The manuscript is then resubmitted to a journal with a lower impact factors and a somewhat less prestigious reputation. This continues until the manuscript is accepted for publication. An open submission process would require authors who resubmit papers to include the original reviewers’ comments. This provides subsequent reviewers and editors with a good sense of what has been said before, what changes have been requested, and allows them the opportunity to build on these comments when making their own. In effect, such a policy would allow reviewers to learn

\textsuperscript{53} Ibid., 972.
from and actively engage with fellow reviewers. This would have an additional practical benefit of expediting the review process, which can otherwise drag on for months. Open submission could contribute to a greater sense of community critical evaluation and collaboration. Alternatively, it might condone laziness in reviewing (“I will just agree with what the other person said”) or lead to greater deference to authority (“If a top medical reviewer thinks this is not good, he or she must be correct”). However, in spite of these potential difficulties, the idea is still a good one in terms of increasing avenues for critical discussion, since it opens up greater discussion among reviewers.

6.1.8. Public Funding of Research

Medical researchers aim to provide accurate information about the value of particular medical treatments. Pharmaceutical companies, in virtue of their status as corporations, are committed to maximizing profits for their stakeholders. Intellectual and financial interests are not always in accord. While these two goals can overlap in some instances, to varying degrees, there are many opportunities for tension and conflict. Steven Lewis describes the partnership between scientific researchers (and the academic institutions in which they work) and pharmaceutical companies as one in which researchers are forced to “dance with the porcupine”. Precautions must be taken to protect the integrity of the dancers if they are to remain free of punctures and lacerations. More outspoken critics have suggested that the appropriate analogy is, rather, one of swimming with the sharks: “those who swim with the sharks may find that they have become little more than shark bait. To avoid such a cruel fate…it might be necessary to decline the swimming invitation altogether.” Commercial interests influence the motives and goals of researchers and, despite self-righteous claims to the contrary, have an impact on the type of research that gets done, how it gets done and who it benefits. Patentable and profitable treatments will always have top priority in the research programs of those interested in making money. While many different concerns are raised about potential conflicts of interest arising out of partnerships between academic researchers and commercial industries, I will focus on the potential for such a partnership – however otherwise beneficial – to diminish the avenues for criticism available to

54 Lewis et al., 783.
55 Shafer, 9.
researchers. Even if confidentiality clauses are disallowed and ghostwriters are restricted by publication rules set by journal editors, commercial interests can limit the inclination researchers have to critically discuss their work with colleagues:

Critics fear that wide scale commercial funding has already produced an erosion of cooperation and community among biomedical researchers. Instead of an easy sharing of knowledge and reagents, one finds something approximating a quasi-Hobbesian war of each (laboratory) against all (others).\textsuperscript{56}

Commercial interests influence the type and quality of interactions among researchers. Self-interest and ego have always played a role in scientific research: there is no simple altruistic story to be told about the interactions of members of scientific communities. However there are differences between the sorts of interactions expected in largely publicly-funded communities and those in which individual researchers, or teams of researchers, are funded competitively by different commercial interests: “[T]here is a growing body of evidence to support the conclusion that the traditional norms of science ought to be placed, forthwith, on the Endangered Species List.”\textsuperscript{57} Concerns about subtle changes to the practice of science were raised in the era of Big Science following WWII, but the impact on clinical trials is more recent. Instead of building on the research of others, researchers have to be concerned that fellow researchers will steal their ideas or beat them to establishing patent rights. These commercial interests influence the interactions between researchers, and while they do not physically limit avenues for critical discussion they make collaborative sharing of ideas less likely. In order to address this problem, researchers convinced of the value of commercial intervention in medical research are limited to the reformist measures outlined above (refusing to sign non-disclosure clauses, declaring competing interests in publications, etc.). Researchers more skeptical about the impact of industry on medical research can join those who call for greater public funding of research and greater government regulation of the pharmaceutical industry.

6.1.9. Emphasis on Original Research

\textsuperscript{56} Schafer, 15.
\textsuperscript{57} Ibid., 16.
In order for physicians to be responsive to the critical discussion occurring in their midst, they have to be attentive to original research. The EBM movement originally emphasized the importance of critical evaluation of research evidence by individual physicians. As we saw in Chapter Four, this requirement has now shifted. Physicians are no longer expected to do all the literature searches and critical analyses themselves. Instead, these tasks are performed by specialized experts. The Cochrane Library, POEM (Patient-oriented Evidence that Matters) Database and American College of Physicians (ACP) Journal Club, among other sources, provide analyses of evidence and function as intermediaries between physicians and original clinical research. Clinical guidelines produced by experts are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.”

Individual physicians are expected to turn to the authoritative digests of research produced by experts when determining what the best evidence is for a new treatment. EBM proponent Gordon Guyatt (along with some of his colleagues) has admitted that:

[N]ot all trainees are interested in attaining an advanced level of evidence-based medicine skills. Our trainees’ responses mirror those of British general practitioners, who often used evidence-based summaries generated by others (72%) and evidence-based guidelines or protocols (84%) but who overwhelmingly (95%) believe that “learning the skills of evidence-based medicine” is not the most appropriate method for “moving… to evidence-based medicine.”

The distinction between evidence-based users and evidence-based practitioners is meant to capture the new type of evidence-based physicians, who use guidelines and recommendations produced by others rather than doing literature searches and critical analyses themselves. The task of critical analysis is contracted out to experts, rather than performed by the individual. This limits the number of different perspectives that are brought to bear on the results of research. While groups that compile evidence involve a variety of reviewers, this is not the same as having individual physicians perform critical analyses for themselves. Reliance on guidelines means that

---

58 Field and Lohr. Timmermans and Mauck also point out: “So many parties have jumped on the EBM bandwagon and so many clinical practice guidelines are churned out by individuals, professional organizations, insurers, and others that the benefits of uniformity may disappear in the cacophony of overlapping, conflicting, and poorly constructed guidelines. With more than 1,000 guidelines created annually, calls for ‘guidelines for clinical guidelines’ have been issued.” See for instance articles by Jackson and Feder; Rosser, Davis, and Gilbart.

59 Guyatt et al., “Practitioners of Evidence-based Care,” 954.
physicians are unable to contextualize evidence in order to determine possible biases and flaws in that research. Moreover, the reviewers themselves introduce a whole new level of biases. A reviewer with a strong bias toward a particular treatment, for instance, could cast the results of a meta-analysis in more favourable light. And all of this is hidden behind an apparently-objective and authoritative guideline. The presence of pre-digested guidelines goes against the original spirit of EBM, and worse, threatens to limit the opportunities for critical discussion on the results of research. The critical analysis at the heart of the original EBM movement is arguably one feature worth retaining – and yet it was the first to be jettisoned. While the difficulties of keeping up with the results of original research are not to be underestimated, reliance on guidelines produced by others limits the opportunity for critical engagement with research. Physicians need to resist the temptation to become evidence users and engage with original research rather than guidelines whenever possible.

6.1.10. Medical Teams

EBM focuses on the decisions made by physicians in the care of patients. In Chapter Four I argued that this neglects the important role played by patients in clinical decision-making. It also neglects the role of the patient’s family and loved ones. Building on this argument, I also argue that it neglects the role played by allied health professionals. EBM is concerned primarily with the decisions made by physicians about patients. But medical decisions are not the exclusive purview of physicians. Nurses, physiotherapists, pharmacists, a variety of medical specialists, ethicists, and even some CAM practitioners have been known to consult together on medical decisions. The physician is not the solitary autonomous decision-maker – even when supported by the latest research evidence. Neglecting the contributions of fellow health care workers impoverishes the EBM model. And requiring that each contributing health professional be evidence-based in their own right does nothing to change the isolated and ultimately autonomous practice of EBM in each particular instance. Further, it fails to require that the evidence retrieved by each health professional be shared and collectively debated. Greater attention to the realities of decision-making in medicine allows us to recognize that physicians are not isolated autonomous decision-makers. Health care teams, often consisting of members with diverse backgrounds, are in a better position to produce interactively objective assessments of evidence.
Thus health care teams play an important role in increase the avenues for critical reflection on evidence at the level of clinical decision-making.

6.1.11. Analytic Training in Medical Education

Physicians are highly trained in the technical aspects of medical research and practice. Many critics have drawn attention to the ways in which this technical focus precludes development of empathetic skills such as good communication and caring. This deficiency is common and is reflected in standards of evidence: in Chapter Five I discussed some of the ways in which qualitative research is unfairly denigrated in the medical context even though it is better suited to investigating the caring side of medicine. The education of physicians is similarly deficient in teaching analytic skills. As the editor of The Lancet, Richard Horton, puts it, “The skill that physicians lack above all is the ability to reason successfully.” Reasoning skills include the ability to critically interrogate an argument, determine gaps in logic, identify fallacies and come to a logical conclusion on whether an argument is sound or, in the case of an inductive argument, cogent. Reasoning is not “the skill of switching on a computer, typing in a few key words and printing out several abstracts of randomized trials or systematic reviews.” While advocates of EBM stressed the importance of understanding rules of evidence (i.e. the rules of the hierarchy) and basic statistics, without broad training in critical reasoning these elements of evidence-based practice remain unsupported. It is no surprise that, in the interim, physicians have turned to guidelines and meta-analyses prepared by others. Greater attention to the skills of critical reasoning of the sort taught in undergraduate philosophy courses – though ideally with a specifically medical focus, and using medical examples – would be a great improvement to medical education and would increase the likelihood that critical debate will occur among physicians and researchers.

7. Summary

---

61 Ibid., 245-6.
As Helen Longino has pointed out, debates over the objectivity of various scientific practices, theories, and methodologies have often proceeded without careful attention to the concept of objectivity, which too often remains “insufficiently examined, a closed box hurled back and forth between rhetorical contestants.” In this chapter I hope to have alleviated some of this mystery and offered some reason to think that the current over-emphasis on procedural objectivity in the medical research context is problematic. I argued that the CCE conception of interactive objectivity has much to add to this discussion. Finally, I drew upon the CCE norms to propose practical mechanisms that lend support to the norms of interactive objectivity in the medical context. In particular, I suggested that various elements of the open science movement could support the interactive objectivity of medical evidence.

Conclusion

Standards of medical evidence are developed on the basis of background assumptions. These assumptions require scrutiny. A critical contextual empiricist (CCE) analysis of evidence-based medicine suggests that assumptions about the need for a more scientific approach to medical decision-making rely on simplistic accounts of science and pay insufficient attention to the cognitive goals and practical aims of medical practice. Assumptions that clinical research evidence should be the base of medical decision-making conflict with persuasive arguments for methodological pluralism in medicine. In addition, the particular assumptions underlying the evidence hierarchy are unjustified; moreover, any attempt to hierarchically organize research methods radically misunderstands the purpose of scientific research. Drawing upon my proposed version of CCE, I criticize the attempt to homogenize medicine which is implicit in current demands that complementary and alternative medical researchers adhere to the rules of evidence developed by EBM. Further, I identify a reliance on procedural accounts of objectivity in EBM, and specify the costs of this narrow focus on research methods which severely restrict the community’s ability to achieve more robust interactive objectivity.

However, a comprehensive appreciation for the limitations of EBM does not leave us without guidance in medicine. In response to an article by Sackett and Oxman, in which the authors trace the biases that can be introduced at various stages of research, physician Jan Brozek writes:

I cannot resist an impression that the EBM as we know it, or at least its part - critical appraisal of the published medical literature, has reached its limits and found itself in desperate need to develop new tools that would allow assessment of the validity of medical research...the ‘usual tools’ like CONSORT criteria, Jadad score, the ITT, completeness of follow-up, concealment of randomization, etc. are no longer sufficient to evaluate the validity of the conclusions accompanying the reports from research. One wonders if maybe it is time to think of developing additional criteria that would help us – users of medical literature – to tell [the] more valid from the less valid conclusions? One may also wonder if it is possible at all since expecting that the investigators/sponsors would give out the necessary information and remain entirely objective in the assessment of their own efforts is naïve.¹

¹Brozek.
Brozeck is correct in identifying the limitations of the EBM approach. It is naïve to expect that the results of research evidence can be summarized, digested and disseminated as “facts” unproblematically. It is also naïve to think that an approach that minimizes the degree of critical interaction with research evidence will remain unsullied by the biases introduced by commercial and other contextual social interests. I argue that greater attention to the revised CCE norms isolates the flaws in these assumptions and, crucially, suggests a way forward. According to my analysis there should be changes to the ways in which research is funded, designed, assessed and distributed and the way medical practice is conceived, including: elimination of confidentiality agreements signed by researchers, disclosure of competing interests in publications, prohibition on ghost authorships, public registration of research, open peer review, open access to publications, open submission processes, increased public funding of research, less reliance on guidelines, greater awareness of the collective nature of medical decisions and, most importantly, greater analytic training for medical professionals. The structural features of the medical community that uphold principles of diversity, transparency, and critical interaction need to be protected and promoted in order to ensure the community retains its ability to produce knowledge. EBM is not just problematic because the particular assumptions upon which it is based are wrongheaded. It is problematic because its proponents were not forthcoming with justifications for these assumptions and for the standards of evidence proposed. Furthermore it is still not clear what would be required to change those standards and critical discussion on the standards has not been encouraged or thoughtfully engaged despite the availability of some appropriate venues for discussion within medicine.

1. Remaining Problems and Future Research

The analysis provided in this dissertation certainly does not solve all of the problems related to evidence and medical decision-making. This is not a surprise given the complexity of the problems. In what follows, I briefly outline a few key problems that remain as well as directions for future research.
1.1. Uncertainty

The uncertainty inherent in medical research and practice cannot be circumvented with rules, guidelines or hierarchies. And it will not be overcome even with the mechanisms of interactive objectivity just outlined. And yet embracing uncertainty poses a significant psychological challenge for those who must make life and death decisions about the care of fellow human beings. There is a tension between the inevitable uncertainty of most clinical decisions and the desire of health professionals to be certain about their treatments. Under these conditions, promises of certainty will always be appealing but must be resisted. As we saw in Chapter Five, Valerie Miké stresses the importance of two basic ethical principles of medical practice in her “ethics of evidence”: make use of best evidence and retain a healthy appreciation for uncertainty at all times. Uncertainty will not be diminished through false claims to certainty but through greater awareness of the limits of evidence. It is to this project that the medical and philosophical communities should turn.

1.2. Quantity of Publications

One of the many motivations for EBM was the increase in the number of biomedical publications. Given thousands of medical journals and hundreds of thousands of journal articles published every year, even a highly motivated physician will be unable to stay on top of all of the relevant articles in his or her area of specialization. The appeal of summarized guidelines and digests is understandable in this situation. I have argued that guidelines and authoritative meta-analyses are dangerous because they limit the amount of critical interaction between physicians and the original research data, and because they decrease the ability of physicians to contextualize the results and assess social biases. This leaves physicians in a difficult situation. After all, there is no way to keep on top of the vast quantity of research being produced every day. The challenges can be diminished somewhat through involvement of the health care team and active involvement of the patient, as discussed in Chapter Seven, but the problem will not, and cannot, be entirely overcome.
1.3. Tendency to Prefer Rules over Reasoning

The original EBM working group said it best: “People like quick and easy answers. Cookbook medicine has its appeal. Critical appraisal involves additional time and effort, and may be perceived as inefficient and distracting from the real goal (to provide optimal care for patients).”

Even with greater training in reasoning and argumentation, many health care professionals will not be interested in doing all of the tedious thinking and critical evaluating required for good medical practice. It will sometimes be appealing to follow simplistic rules and pre-packaged guidelines. Beyond the requirement that physicians gain an appreciation for the basic principles of reasoning through their medical education, I am not sure what can be done to change this attitude.

1.4. Ethics and Epistemology

Epistemological assumptions have an effect on the ethics of medical practice. For instance, a great deal of attention has been paid to the concept of informed consent in bioethics. In the bioethical literature there are hundreds of articles offering careful articulations of the elements of informed consent (usually including disclosure, understanding, voluntariness and competence), the goal of informed consent (respect for autonomy and self-determination), as well as thoughtful articulations of different decision-making models and guidelines for the physician-patient relationship. The question too often glossed over in these otherwise thoughtful discussions is where the information comes from. There is a tendency to neglect the nature of the information in informed consent. The assumption is that the facts are unproblematic. The focus is on how best to convey the evidence, not on what evidence to convey or how to decide which evidence is worth conveying. Epistemological questions about the justifications of different sources of evidence come into play as physicians choose which information is relevant to their patients. These epistemological choices thus affect the quality of patients’ informed consent. As a result, epistemological claims have ethical import in the domain of medical practice. The “ethics of evidence” is an important subject and one that has received scant attention thus far in the

---

2 Evidence Based Medicine Working Group, 2423.
bioethical literature. It is imperative that research projects are conducted on the ethics of medical evidence.

2. Themes

There are three themes that run throughout this thesis. The first theme is the importance of transparency: the need to be upfront about values. I provide empirical evidence of bias in medical research and defend an account of interactive objectivity. Throughout, the emphasis is on the pervasive presence of values in knowledge-production. It is dangerous to pretend we have superseded these values by projecting a false image of objectivity. We have empirical, ethical, and epistemological reasons to value diversity, transparency and critical interaction.

The second theme is the importance of community. Scientific activity requires the activity of more than one individual. Medical research is no exception. Failure to attend to better and worse ways of organizing scientific communities can leave them open to a variety of abuses. A comprehensive epistemology of medicine will include specific, practical and achievable criteria for keeping communities sufficiently open. Relying on mechanisms that apply only to individuals limits our ability to protect communities from the idiosyncratic views of the powerful minority.

The third theme is the importance of diversity, creativity and innovation. Whether it is the methodologies and theoretical frameworks proposed by researchers in complementary and alternative medicine or the diverse perspectives of philosophers such as myself addressing issues in the domain of medicine, diverse perspectives allow for new critiques and novel proposals for change and improvement. Interdisciplinary research and even research that brings together a variety of sub-disciplines within one discipline (such as bioethics, philosophy of medicine and social epistemology in this dissertation) provides vital new perspectives and fruitful new areas of research. This theme is reinforced in both the content and structure of this project.

3. Final Words

In the Introduction to this dissertation I stated that I intended to produce a critical analysis of current standards of evidence that was as charitable to EBM as possible. To some readers it may
appear as though I have broken this promise. I do, after all, make a number of strong claims about the failings of EBM throughout the dissertation. In chapters four and five, in particular, I argue that the fundamental assumptions of the movement are unjustified and, in some cases, just plain wrong. I suggest that EBM assigns unwarranted creedal value to the products of particular research methods. This comprehensive critical analysis does not, however, indicate a lack of appreciation for the motivations behind EBM. I have a great deal of respect for the physicians who first proposed EBM. Rather than complacently accepting a system rife with appeals to authority and too slow to give up old, often harmful, habits, the members of the EBM working group sought to bring about a more rational, more rigorous and more humane medical practice. While I think the details of their attempt to improve medicine were less than ideal, I appreciate that EBM has forced physicians to talk about standards of evidence, the elements of clinical decision-making and methods of assessing clinical research. And while the movement has shifted in recent years, for the short time that it emphasized critical thinking I think it provided an important perspective on the value of analytic skills for medical professionals. Further, when you look at the prominent physicians today who advocate for improvements to medicine, the old guard of EBM are notable. For instance, it is Gordon Guyatt who drew attention to the value of the otherwise little-known n-of-1 method in research.³ It is members of the Cochrane collaboration who have most actively lobbied for a clinical trials registry.⁴ And it is David Sackett and colleagues who have written the most comprehensive and provocative guide to the ways in which research evidence can be biased by corporate interests.⁵

Medical decisions are morally weighty: they are fraught with potential harms to innocent patients. The weight of these decisions demands that physicians justify their decisions to the highest standards. Appeals to authority are inadequate, and evidence – whether from clinical research or patients’ own experiences – should be an important factor in decision-making. That is not to say, however, that clinical research evidence should be the base of medical decision-making, nor that certain research methods, all else being equal, provide inherently better evidence than others. And neither medical research nor medical practice will progress as long as we leave these false assumptions in place.

³ Guyatt et al., “Determining Optimal Therapy: Randomized Trials in Individual Patients.”
⁴ Rennie.
⁵ Sackett and Oxman. Sackett also credits Iain Chalmers in the article.
Evidence-based medicine is a term “both evocative and trite.” I have been astonished by the rhetorical power of evidence-based practice, the language of which has spread far beyond the realm of health care. Even in the face of extensive critique, this movement has gained momentum all over the world. This is due in no small part to the easy marketability of the name, but it is more than this. People want to rely on facts, on objective truths, when making decisions. We all rest easier knowing that physicians have managed to beat back uncertainty – even if only for a few moments until the next study comes out. The practice of medicine is challenging enough without second-guessing everything you do. But what is the price we pay for setting up seemingly objective procedures, hierarchies of research methods and rigid rules of practice? What is the price of false generalizability, false certainty, false objectivity?

I think there are those who would argue that the price, however high, is worth paying because it is “better than nothing” and certainly a lot better than the paralysis of uncertainty. I disagree. I do not think we progress in science, or any knowledge-seeking endeavour, when we settle for alleged certainty over accuracy. We do not advance medical research by limiting the scope of what is investigated to that which fits the narrow received view of good science. We do not advance medical practice by requiring that physicians make decisions according to clinical decision-making rules that hide important, and often questionable, background assumptions. We do not advance human health by insisting that there is one correct way to practice medicine or one best method for investigating the complex, dynamic and often unpredictable facets of disease, illness and suffering. I have argued that the standards of evidence currently accepted in the medical domain serve to diminish the objectivity, transparency and diversity of medical research and, by extension, medical practice. As a result, EBM diminishes the chances for innovation in medicine and negatively impacts the health and well-being of patients.

Although this discussion has focused on medical standards of evidence, similar issues are present in such diverse fields as agriculture, ecosystem management and education. Whenever people are tempted to apply procedural mechanisms rather than critical reasoning or whenever they are inclined to prefer a false sense of certainty over a true sense of uncertainty, there will be a market for evidence-based practice in its various guises. I join those in the medical community who lament the intellectual laziness behind these choices, and who are outraged at the resulting harm to patients. The social institutions of science – publication, funding, education, etc. – can

---

6 Daly, 11
be arranged in better and worse ways in terms of their ability to address this problem. Only by paying attention to the epistemological weight of these social organizations and practices will we be able to address the most significant problems facing medical research and practice today. And only then will we be prepared to defend these social institutions and arrangements against encroachment by commercial and other social interests.
Bibliography


Canadian College of Naturopathic Medicine (CCNM) Website. www.ccnm.edu/about.html.


http://www.compmed.umm.edu/Cochrane/cam_reviews.html.


CONSORT (Consolidated Standards of Reporting Trials) Group. “CONSORT Statement.”
http://www.consort-statement.org/.

http://education.guardian.co.uk/higher/profile/story/0,,1698284,00.html.


Guyatt, G.H. et al. “Practitioners of Evidence Based Care. Not all clinicians need to appraise evidence from scratch but all need some skills.” British Medical Journal 320 (2000): 954-5.


Guyatt G.H., D.L. Sackett, and D.J. Cook [Evidence-based Medicine Working Group]. “Users’ guides to medical literature. II. How to use an article about therapy or prevention. B. What were the results and how will they help me in caring for my patients?” The Journal of the American Medical Association (1994) 271 1: 56-63.


Haack, S. “Trial and Error: The Supreme Court’s Philosophy of Science.” American Journal of


MYCaW Website: http://www.bristol.ac.uk/hsrc/research/other/mymop/mycaw/index.html.

MYMOP Website: http://www.bristol.ac.uk/hsrc/research/other/mymop/index.html.


Pittler M.H. and E. Ernst. “Horse chestnut seed extract for chronic venous insufficiency.”


Quackwatch Website: www.quackwatch.org.


Sandelowski, M. “‘To Be Of Use’: Enhancing the Utility of Qualitative Research.” *Nursing Outlook* 45 (1997): 125-32.


