RICE UNIVERSITY

What is Evidence-Based Medicine?:
Defining and Critiquing the Theory and Practice of Evidence-Based Medicine

by

Jennifer A. Bulcock

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

H. Tristram Engelhardt, Jr.,
Professor of Philosophy

Richard E. Grandy,
McManis Professor of Philosophy

Melinda B. Fagan,
Associate Professor of Philosophy,
University of Utah

Rachel Kimbro,
Associate Professor of Sociology

HOUSTON, TX
DECEMBER 2015
ABSTRACT

What is Evidence-Based Medicine?:
Defining and Critiquing the Theory and Practice of Evidence-Based Medicine

Jennifer A. Bulcock

Over the past 23 years evidence-based medicine (EBM) has been incorporated into the healthcare systems of the United States, the Netherlands, England and many other countries. Despite its widespread acceptance, foundational questions persist regarding what EBM is and whether its foundational assumptions are justified.

It is the contention of this project that EBM cannot be coherently understood as a singular, unified theory of medicine. Rather, it is most appropriately conceived of as a cluster concept or brand name under which several theories are constructed and are characterized to varying degrees by EBM’s four primary principles: 1) randomized controlled trials (RCTs) are the gold standard research methodology, 2) clinicians should distrust authorities, 3) clinical experience and clinical expertise are not high quality sources of evidence and 4) the standardization of medical practice is an ideal. Particular focus is given to EBM’s account of clinical practice and its approach to assessing research evidence of clinical effectiveness.

EBM’s account of clinical practice and its historical genesis are expounded and analyzed. It is argued that EBM’s initial articulation of a theory of clinical practice had to be abandoned because it was impossible to realize in practice. Further, given its most recent articulation, it is maintained that EBM no longer constitutes a unique theory of clinical practice because it permits too much variance.

EBM’s approach to assessing research evidence of clinical effectiveness is also carefully analyzed. It is argued that EBM’s hierarchy of evidence is unjustifiably founded on the
epistemological assumptions that 1) RCTs have a superior ability to identify causal relationships between treatments and their effects and 2) RCTs produce less biased results. Additionally, it is argued that the RCT’s limited external validity causes further epistemic and practical problems for the evidence hierarchy. Accordingly, it is argued that the current categorical interpretation of the evidence hierarchy is epistemically unjustifiable and should be abandoned.
ACKNOWLEDGEMENTS

This dissertation took many years of long, hard work. That work was not mine alone. At the completion of this project it is overwhelming to look back on my time in graduate school and college and acknowledge everyone who helped get me to this point. I have been very fortunate throughout my academic career to have exceptional mentors, intelligent and challenging peers and colleagues, truly generous and kind friends, and a loving and supportive family. This accomplishment is as much theirs as it is mine. There are no words sufficient to express my gratitude.

There is one person to whom I owe my academic career and my love of philosophy. If it weren’t for Nick Smith, my advisor and mentor during my undergraduate and graduate years at the University of New Hampshire, I may never have found my true calling in life: writing and teaching philosophy. Not only did Nick introduce me to philosophy, he taught me how to write, how to think, and how to argue. Nick literally read, discussed, and commented on hundreds of pages of my writing during my years at UNH, including seven classes worth of papers, an undergraduate honors thesis, and a Master’s thesis. It goes without saying that without Nick’s mentorship my academic trajectory would’ve been much different.

While Nick helped me firmly set my course down the rabbit hole of academic philosophy, my graduate advisor, H. Tristram Engelhardt, Jr., deserves my deepest gratitude. I struggle to find the words to even begin to enumerate the many debts, both academic and personal, I owe to Tris. Over the past eight years he has been a constant mentor and advocate. From him I have learned so many lessons, some found in the pages of philosophy texts and others far beyond the pages of any book. Tris has pushed me far beyond the limits I saw for myself and has made me a better philosopher, teacher, and person because of it. I am deeply
grateful for Tris’ devotion to his students, the time and energy he has invested in my academic
and professional development, and, of course, for his patience and guidance in bringing this
project to completion.

I also owe a great deal of gratitude to my other committee members. If it weren’t for
Melinda Fagan I would almost certainly be writing the acknowledgements for an entirely
different dissertation. It was while auditing Melinda’s philosophy of science course that I came
across the article by Nancy Cartwright that sparked my interest in evidence-based medicine and
eventuated the writing of this dissertation. Melinda has been a constant source of encouragement
and support both for my research and during my time on the academic job market. I am
exceptionally grateful to Melinda for the many conversations we have had about the philosophy
of science and, in particular, mechanisms in medicine.

Richard Grandy has been the voice of reason for me during my time in graduate school. It
is not an overstatement to say he has been integral in keeping me sane over the past eight years.
Of all the professors at Rice I think I have knocked on his door and sat in his office the most. He
has given me honest and wise advice at critical times in my academic and professional
development and I am truly grateful for his patience, kindness, and mentorship.

I am grateful to my outside reader, Rachel Kimbro, for signing on to my project in its
final days and helping me to bring the project to completion. I also want to thank the faculty of
Rice University’s philosophy department for many engaging classes, invaluable discussions, and
generous support throughout my years in the department. Minranda Robinson-Davis has also
been integral to my success in graduate school and I thank her for all the work she does that goes
unnoticed, many good conversations, and most of all her patience and kindness.
While my academic debts are great, my personal debts are even greater. As any graduate student knows, your friends are often what keep you afloat after devastating comments on a manuscript, weeks of writing with little sleep and much stress, and those final days before taking your qualifying exams or defending your dissertation. In graduate school I have found friendships I suspect will last a lifetime. I do not have the space here to thank each of my friends and peers individually, but I can say that each of them has played a very important part in my graduate school experience both intellectually and personally. The love, kindness, support, and generosity I have received from each of them is deeply appreciated and I hope reciprocated. I am honored to have such amazing people among my closest friends. These wonderful people include Heather Phillips, Joe Adams, Sherry Ya-Yun Kao, Julia Bursten, Carissa Phillips-Garrett, Amanda Bancroft, Graham Valenta, Brandon Williams, and Ruth Williams.

Lastly, I want to thank my family for their unconditional love, understanding, and support. In many ways I would never have made it through college, graduate school, and the job market without the support of my parents. My father, Thomas Bulcock, has always been my biggest cheerleader and has believed in and encouraged me every step of the way. My mother, Judith Cassidy, has always been a source of support, fielding many panicked and irrational phone calls over the years and talking through many of life’s most important decisions. I love you both so much. You’ve both always been my rock and I couldn’t have done any of this without you.

And to my furry little monsters, Ayla and Tucker, who have kept me sane and made me laugh when I’ve wanted to cry, I owe you many walks and belly rubs.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>1</td>
</tr>
<tr>
<td>I. What is Evidence-Based Medicine?</td>
<td>3</td>
</tr>
<tr>
<td>History, Motivations, and Identity</td>
<td></td>
</tr>
<tr>
<td>II. A Theory of Clinical Practice: Definitions &amp; Critique</td>
<td>31</td>
</tr>
<tr>
<td>III. Evidence-Based Medicine in Practice</td>
<td>59</td>
</tr>
<tr>
<td>IV. Assessing Research Evidence of Clinical Effectiveness: The Evidence Hierarchy</td>
<td>94</td>
</tr>
<tr>
<td>Bibliography</td>
<td>116</td>
</tr>
<tr>
<td>Appendix A: Mechanistic Evidence in Evidence-Based Medicine</td>
<td>124</td>
</tr>
<tr>
<td>Analyzing the Russo-Williamson Thesis</td>
<td></td>
</tr>
<tr>
<td>Appendix B: Descriptions of Methodologies</td>
<td>141</td>
</tr>
</tbody>
</table>
PREFACE

This dissertation is a practical and theoretical project locating itself within the field of the philosophy of medicine, understood as a subfield of the philosophy of science. The analyses undertaken advance a foundational critique of evidence-based medicine as an account of clinical practice, an account of clinical decision-making, and as an account of how to assess research evidence of clinical effectiveness – all of which have significant theoretical and practical consequences for an evidence-based theory of medicine. Because the EBM literature lacks precise definitions of any kind, fails to provide a coherent foundational theory,¹ and seldom explicitly recognizes that EBM is not a singular, unified theory of medicine,² preliminary work is done to define and delineate the theories of EBM analyzed in the remainder of the project.

The analyses undertaken in this dissertation have practical, theoretical, and philosophical importance. Of primary philosophical importance is determining whether and to what extent first principles can be articulated for evidence-based medicine as well as whether a coherent epistemological theory is employed. It is well acknowledged that EBM has developed without careful attention to first principles, precise definitions, and sound argumentation - the bread and butter of philosophical analysis. As Andrew Miles describes it, EBM was “a systematic attempt…to revolutionize the whole nature of medical practice in the absence of a soundly...

worked out theoretical base and without any form of general acceptance of EBM principles by
the international medical community.” That is, EBM provided seemingly intuitive principles,
definitions, and guidelines without providing a rigorous theoretical basis for them or their
implementation. Similarly, Bruce Charlton explains,

EBM sprang from the womb fully formed as a self-evident truth. EBM was not a hypothesis but a circular and self justifying revelation in which definition supported analysis which supported definition – all rolled-up in an urgent moral imperative. Therefore, EBM was immune to the usual feedback and critique mechanisms of science; EBM was not merely disproof-proof but was actually virtuous – and failure to acknowledge the virtuous authority of EBM and adopt it immediately was not just stupid but wicked.

Some critics believe this has led to EBM functioning as no more than a “brand name” describing different accounts of medical practice, education, and methods for evaluating clinical research evidence. The analyses of this dissertation will reinforce the veracity of this claim.

Taking all of this into consideration, it is clear that EBM stands to benefit greatly from sustained philosophical critique aimed at delineating its foundational tenets and its most reasonable theoretical construction in addition to defining its foundational concepts. Such work is invaluable to the philosophy of medicine, in particular medical epistemology, because, having unjustifiably been set forth as a new paradigm of medicine, EBM has caused great confusion regarding what evidence clinicians should judge epistemically valuable and how different types of evidence should be incorporated into clinical decision-making. It is the aim of this dissertation to add philosophical rigor and nuance to evidence-based medicine and to provide answers to long standing questions in the interest of positively impacting or, at the very least, clarifying evidence-based practice.

---

4 Charlton, “Zombie Science,” 932.
CHAPTER 1
WHAT IS EVIDENCE-BASED MEDICINE? HISTORY, MOTIVATIONS, AND
IDENTITY

I. Introduction: A Very Brief History

Over the past 23 years Evidence-Based Medicine (EBM) has been incorporated into the
healthcare systems of the United States, the Netherlands, Australia, England and many other
countries in the developed world. In its initial articulation in 1992, EBM was characterized in
the medical literature as a “new paradigm” for medical practice that “de-emphasiz[ed] intuition,
unsystematic clinical experience, and pathophysiological rationale as sufficient grounds for
clinical decision making and stress[ed] the examination of evidence from clinical research.”
The Evidence-Based Medicine Working Group (EBMWG), which characterized EBM in this way,
found at the time that medical practice was changing in such a way that clinicians needed more
effectively to use the medical literature in guiding their medical practice. They asserted that it
was necessary for one to “move from basing medical decisions on the “unsystematic” judgment
of an individual clinician, based on experience or the findings of the bench or basic sciences, to

---

1 The term “evidence-based medicine” actually appeared in Gordon Guyatt’s 1991 editorial entitled, “Evidence-Based Medicine,” in the *ACP Journal Club* (114: A16). However, the EBMWG’s 1992 article is oft cited as the initial presentation of the concept because it formally advances EBM as a new paradigm of medical practice and makes some attempt at fleshing out its tenets. Guyatt’s 1991 essay foreshadows a new “way of the future” that involves the “inclusion of new evidence into clinical practice,” but does not formally present EBM as a new concept or theory.


4 Ibid., 2420.
the more “systematic” and “relevant” outcomes of patient-related clinical research [randomized controlled trials (RCTs)].” The rapid development of, demand for, and use of clinical trials to assess the effectiveness of treatments are what the EBMWG understood as the primary catalyst for this shift.

To provide an accurate picture of the increase in availability of RCTs, consider the fact that in 1955 “there were only 347 reports of randomized trials. Half a century later, about 50,000 reports of randomized trials were being published every year, with the total number of trial reports by then exceeding half a million.” Accordingly, it goes without saying that RCTs have a dominant presence in the medical literature and have for some time, simultaneously creating and responding to the need for additional consideration of RCT evidence. As S.G. Henry, R.M. Zaner, and R.S. Dittus have put it, “the lynchpin of the evidence-based medicine movement…is the premise that the results of well-designed studies, especially randomized control trials should govern medical decisions.” Consequently, the privileging of RCTs within EBM has been a main tenet of its approach to medicine as well as the focus of much discussion and debate.

While the rapid proliferation of RCTs has provided clinicians with previously unavailable information about the safety and efficacy of interventions, the value of which EBM has recognized and encapsulated in its approach, the motivations for pursuing this new practice of medicine were somewhat diverse. In general, the original impetus for developing the EBM movement was reportedly a concern for how to teach and practice medicine more effectively.

---

8 Evidence Based Medicine Working Group, “Evidence-Based Medicine,” 2420-5.
Four well recognized motivations\textsuperscript{9} said to be driving clinicians’ dissatisfaction with the practice of medicine during this time include: 1) a desire to eliminate, or at least minimize, the inverse relationship between healthcare costs and patient outcomes (i.e., a desire for the number of healthcare dollars spent and the positive patient outcomes gained to be positively correlated); 2) a recognition that the dogmatic teaching practices of senior medical faculty in the clinic sometimes led to inaccurate and/or outdated recommendations for treatment being passed down without challenge, oftentimes leading to patient harm; 3) a growing dissatisfaction with “expert consensus”\textsuperscript{10} as the primary method for generating evidence on the safety and efficacy of medical interventions and devices; and 4) a desire to achieve greater standardization in medical practice\textsuperscript{11} such that two clinicians presented with similar patients would reach similar judgments regarding the appropriate diagnosis and treatment.

In the remainder of this chapter the four motivations for EBM mentioned above will be further fleshed out and some initial analyses provided. Such explanation and analyses will lay a better foundation for understanding EBM, its historical development, its primary tenets, and the criticisms it has faced. Examining EBM’s motivations will allow for the extraction of an evaluative framework by which one can measure the success of EBM in practice as well as the coherency of its many definitions over time. Once this foundation has been laid, an initial answer to the important and complex question, “what is EBM?” will be provided. By considering this

\textsuperscript{9} While it is the case that the following four motivations have been set forth in an overwhelming number of contributions to the literature, there are a few scholars who deny the existence of these concerns in the general medical community.

\textsuperscript{10} Expert consensus is reached in a manner intuitive to the name. Expert consensus panels are comprised of leading experts in a particular specialization or subspecialization (e.g., pediatric oncologists, dermatologists, internists, etc.) for the purpose of reaching a consensus about a particular medical fact (e.g., defining a new disease or syndrome; determining whether a particular intervention was effective; determining best practices for treating a particular disease or illness).

\textsuperscript{11} EBM’s aim of standardizing medical practice has been criticized for leading clinical practice away from the goal of individualizing patient care. For further discussion of this criticism see: Romana, H.W., “Is Evidence-Based Medicine Patient Centered and Is Patient Care Evidence Based?,” \textit{Health Serv Res} 41(2006): 1-8.
question, an argument will be made against its being understood as a singular, unified theory of medicine and in favor of EBM being considered a “brand name,” in line with the arguments of Michael Loughlin and Andrew Miles. Ultimately, it will be argued that EBM admits of at least three separate theories – a theory of clinical practice, a theory of medical decision-making, and a theory for assessing research evidence of clinical effectiveness – that need to be understood and analyzed separately. The claim that these three theories actually constitute different parts of a unified theory of medicine will be rejected pointing to the lack of necessary conceptual links among them and the isolation of their relevant literatures from each other as evidence. Additionally, the analyses undertaken in this chapter will provide the necessary foundation and framework for understanding the arguments advanced in the remaining chapters.

II. Motivations: Discontent and the Development of EBM

Examining the motivations underpinning the development of EBM helps to illuminate how and why its theory and practice have developed in the manner they have. Additionally, some of EBM’s problematic tenets (e.g., deemphasizing authority) are rooted in these initial motivations and, therefore, indicate some initial guidelines for analyzing the successful practice (on its own account) of EBM.

II.1 Standardization: PICO(T) Process and Guidelines

The desire for standardization among clinicians arose from the related concerns of dogmatic teaching practices and reliance on the expert opinion of other clinicians. The concern was that if

---

one based treatment decisions on the recommendations of one’s mentors and senior clinicians
then there would necessarily be “high variability” in the recommended treatment from
institution to institution based on the biases and experiences of the particular set of clinicians to
which one was exposed. This is not to suggest that recommendations would be dramatically
different or medically unsubstantiated (although perhaps outdated), but variance in the treatment
protocol would be possible. Having standardized guidelines or a standardized process in which
clinicians would seek out the most current research evidence (available to all clinicians) would
help to ensure that the same patient would receive the same treatment recommendation
regardless of the clinic or emergency room to which she was admitted.14

The desire for standardization among clinicians has been formalized through the
development of evidence-based clinical practice guidelines and the six-step procedure for
evidence-based clinical decision-making, commonly referred to as the PICO or PICOT process
(Population/Patient Problem, Intervention, Comparison, Outcome, Time). Clinical practice
guidelines are evidence-based15 principles meant to guide the practice of clinicians when making
treatment and diagnosis decisions. For example, the National Guideline Clearinghouse provides
the following guideline for the treatment of tuberculosis: “Most patients with pulmonary
tuberculosis not previously treated should be treated using a short, 6-month regimen consisting
of an initial 2-month phase of isoniazid, rifampicin, pyrazinamide and ethambutol and a 4-month

13 D. Mayer, Essential Evidence Based Medicine, second edition. (New York: Cambridge University Press, 2010),
11.
14 It is of course possible that even if the ideal of standardized treatment recommendations were achieved (i.e., every
clinician would prescribe the same treatment based on the available evidence) the same patient would not be
prescribed the same treatment at different clinics due to other factors such as resource limitations of the hospital or
the personal preferences of the patient. However, the concern here is focused on clinicians reaching the same
medical judgment regarding treatment and not the socioeconomic factors that may necessitate recommending a
different and suboptimal treatment.
15 Clinical practice guidelines are not an artifact of EBM. Clinical practice guidelines were recommended as the
result of the consensus conferences prior to and during the beginning years of EBM. Miriam Solomon gives an
excellent historical account of consensus conferences, their resultant guidelines, and the rise of EBM in her book, M.
maintenance phase of isoniazid and rifampicin.” From such guidelines clinicians are given a plan of treatment that all other clinicians faced with similar patients should also be using, thereby providing a standard treatment recommendation. Absent a clinical practice guideline, clinicians should use the PICO/PICOT process to reach an evidence-based judgment. The thought behind the six-step process is: if all clinicians judiciously and carefully follow it, they will, in the end, reach the same clinical judgments. An example of EBM outlined as a six-step process can be found in Dan Mayer’s textbook, *Essential Evidence-Based Medicine, 2nd* Edition. It states that when practicing EBM one should:

1. Craft a clinical question that includes the patient, the intervention, the comparison, the outcome of interest, and the time frame (often called the PICO or PICOT formula).
2. Search the medical literature for those studies that are most likely to give the best evidence. (This step requires using good searching skills using medical informatics.)
3. Find the study that is most able to answer this question. Determine the magnitude and precision of the final results.
4. Perform a critical appraisal of the study to determine the validity of the results. Look for sources of bias that may represent a fatal flaw in the study.
5. Determine how the results will help you in caring for your patient.
6. Evaluate the results of applying the evidence to your patient or patient population.

---

17 The use of the term “studies” here implies that the only trustworthy sources of evidence one may use are clinical trials. The use of this term effectively rules out case studies, clinical expertise, etc. as trustworthy evidence sources.
18 Mayer, *Essential Evidence Based Medicine, 2nd* ed., 14-15. Articulations of the 6-step process vary from author to author but all assert the same six general steps.
19 It is important to note that clinical expertise, patient preferences, pathophysiological rationale, and system (or contextual) features are not mentioned in this process. One may argue that these other sources of knowledge/evidence come into play in step 5. While this claim appears necessarily true, the larger concern/complication is that the PICO/PICOT process does not provide any guidance as to how a clinician is suppose to balance the empirical evidence garnered during the PICO/PICOT process with these other forms of knowledge/evidence.
20 The formal six-step process used currently in textbooks like Mayer’s stems from earlier articulations of a similar process by the EBMWG and its members. The content of the process is similar, however the particular aim of each step has slightly shifted over the past 18 years. David Sackett, in his 1997 essay, sets forth five steps for practicing EBM:
From this six-step process one can see how EBM has sought to standardize medical practice by requiring clinicians to adopt an identical process supported by a common literature base when making clinical judgments. As a result, two clinicians should theoretically reach similar judgments regarding which medical intervention(s) is indicated for similar patients presenting with similar symptoms. If clinicians achieve such standardization, the hope is that patient outcomes will not only be similar, but optimal.

While the six-step process concisely lays out the primary components of practicing EBM, it may not be immediately apparent what each step requires in practice. As such, it is useful to illustrate each step in further detail.

**Step #1:**

The first step instructs the clinician on how to develop the most relevant clinical question for her patient (or patient population) capable of generating accurate and useful research results. This step is fairly straight forward; however, it should be noted that not every clinical question will be answered by the literature and often the literature lacks relevant results for certain patient populations. This is particularly true when dealing with elderly or pediatric populations due to the ethical constraints and/or confounding co-morbidities that prevent such populations from participating in research.

---

1. Convert these information needs into answerable questions;
2. Track down, with maximum efficiency, the best evidence with which to answer them (and making increasing use of secondary sources of the best evidence);
3. Critically appraise the evidence for its validity (closeness to the truth) and usefulness (clinical applicability);
4. Integrate the appraisal with clinical expertise and apply the results in clinical practice; and
5. Evaluate one’s own performance.


21 Here “practicing EBM” has a narrow interpretation meant only to refer to the practice of clinical decision-making - specifically determining what the best research evidence is.
Step #2:
The second step encapsulates the original quintessential component of EBM practice—searching the medical literature for the “best evidence” useful for answering a particular clinical question. Generally, this step requires performing a search on PubMed\textsuperscript{22} using the question constructed in step one.

Step #3
The third step further fleshes out the research process by instructing the clinician to narrow the pool of studies she has found. In particular, one is instructed to find the study that is most able to answer one’s clinical question. There is a lot packed into this step and much is lost to the casual reader (although as a quick and concise outline of the EBM process this is understandable). To understand what is necessary for practicing EBM more needs to be said here. The study one is instructed to find is the one that most closely matches the particular clinical context, condition, and patient one is attempting to address (e.g., tuberculosis in an Asian, pediatric male). In reality, this is no simple task. On the one hand, there are many instances in which one will not find a study exactly addressing the patient population with which one is concerned (e.g., Asian, pediatric males). On the other hand, some cases (such as those concerning common, well-researched diseases such as strep throat or myocardial infarction) will yield an abundance of information with many studies addressing the particular clinical question with which one is concerned.

In the first case, one will have to settle for the best available evidence (e.g., tuberculosis in Asian, adult males) and make extrapolations from that evidence to the patient population one is charged with treating. In the second case, other problems, e.g., having conflicting results, will

\textsuperscript{22} PubMed is the premier search engine for medical research.
make it difficult to determine which trial(s) provide(s) the best evidence, i.e., the most reliable.

At this point one should notice that how “best evidence” is defined shifts depending on the evidence one has available.

Step #4:

In step four one is instructed to conduct a “critical appraisal”\(^\text{23}\) of the study, which requires looking at its methodology to ensure that it is unbiased (i.e., the study was appropriately blinded, etc.) and that the study is reliable (i.e., it has a sufficient sample size, an appropriate control group, etc). For the clinician faced with an abundance of evidence from the research literature, this step may weed out a number of the studies initially judged to be important. For the clinician with too little evidence or with only a few promising studies, this step serves as a quality control measure to make sure the study one is relying on is actually a good quality study free from bias and confounding. Any biased or methodologically flawed study should be discarded from consideration even if it is the only evidence one has.

One issue that should be mentioned at this point, which will be addressed in further detail in chapter three, is that not all clinicians have the time or training in trial methodology and/or statistics to be able to complete this step with ease. As a result, systematic reviews have been introduced as a means of circumventing part of the six-step process while still (hopefully) maintaining standardized judgments.\(^\text{24}\) That is, systematic reviews make steps 3-4 unnecessary for individual clinicians. Even though systematic reviews go a long way in attempting to

\(^{23}\) There are a number of formalized methods for conducting a critical appraisal, including the Critically Appraised Topic, Disease Oriented Evidence, Patient-Oriented Evidence that Matters, and the Journal Club Bank. The details of the variance among these methods are not important for the arguments being made here. In general, critical appraisal is understood to be a critical review of a study or set of studies to ensure the trial methodology was high quality and that biases and confounding were controlled for.

\(^{24}\) Systematic reviews are reviews of the current literature for a particular disease that distills all currently available evidence into a quickly accessible summary, preventing clinicians from having to search the literature themselves. Meta-analyses involve pooling the data from similar RCTs for a particular intervention with the aim of increasing the power of the results by generating the largest possible effect size.
standardize clinical recommendations, it still remains the case that many clinicians fail to consult them or, even when consulting them, reach different conclusions than those set forth in the recommendations.\(^\text{25}\)

**Step #5**

In the fifth step the assumption is made that one has been successful in narrowing one’s results down to a single (or a few) high quality study that is applicable to one’s patient, although most likely not a perfect match. With the evidence one has, the clinician is then tasked with determining how the evidence applies to one’s patient. Depending on how exactly the evidence one has applies to one’s patient this step may require a study to be *externalized* in two different ways.\(^\text{26}\)

If one’s evidence is a close match to one’s patient profile (e.g., the study reports the efficacy of intervention X for treating tuberculosis in Asian, pediatric males and one’s patient is an Asian, pediatric male) then one will only need to consider whether there are clinical factors particular to one’s patient that may affect the efficacy of intervention X for one’s patient. Such factors may include co-morbidities, atypical symptomatology, etc. This step also requires consulting one’s clinical experience and knowledge, although this is not made clear in the presentation of the PICO/PICOT process. If after considering the unique clinical factors of one’s patient one judges intervention X remains the best option for one’s patient, then the clinician should prescribe intervention X.

The second way in which clinicians may need to externalize study results is in a case where the evidence one has is for a patient population other than the population to which one’s patient belongs. For example, if one is only able to locate studies with Caucasian adult males as

\(^{25}\) There are a number of issues involving the use of meta-analyses and systematic reviews relating to expertise, authority, and bias which be discussed in more detail in later chapters.

\(^{26}\) The problems with externalizing study results are further discussed in chapter 4.
the patient population, then one will need to consider how the efficacy of intervention X may change when applied to Asian pediatric males. For instance, one will need to consider how the intervention may interact differently in the body of an adult compared to a child, or how one’s physiology may differ being of Caucasian versus Asian decent. In most cases, interventions tested on adults will need to be adjusted when administered to a child due to differences in average body weight, developmental stages, hormone levels, etc. When dealing with differences in race or ethnicity, differences in physiology may or may not be known, present, or relevant. However, if differences are present, e.g., sensitivity to environmental or dietary factors, they may counteract or complicate the efficacy of an intervention and, therefore, must be taken into account. Once one has made these determinations, one will also need to take into account the individual clinical factors of one’s patient as mentioned above. Again, one’s clinical experience and expertise is essential in determining how to apply evidence to a particular patient.

Step #6:

Once one has made determinations about the applicability of one’s evidence to one’s patient, a clinician should have the requisite information to be able to make a clinical judgment and administer a treatment plan. However, as the sixth and last step indicates, the clinician must evaluate the results of applying the evidence to one’s patient. This step is not only suggesting one should measure how good the treatment is—that is, monitoring one’s patient to make sure the intervention is successful at alleviating the patient’s condition and does not cause unacceptable harm—it is also suggesting one should critically reflect on whether the treatment is efficacious in the ways suggested by one’s evidence, or if there are reasons for doubt or concern. If there are reasons for doubt or concern (i.e., serious adverse reactions or harm or the treatment is inefficacious), there may be reason to report the clinical outcomes of one’s patient and/or
indicate to the medical community that further research needs to be conducted. Once one has reflected on the reliability and usefulness of the evidence, the six-step process is complete and the clinician has successfully engaged in the practice of evidence-based medicine.

At this point, two comments need to be made about the six-step process for practicing EBM. First, as was noted briefly in the explanation of the six-step process, clinical experience and clinical expertise are imperative for the successful practice of EBM, especially when determining how to apply evidence to a patient. This fact is often underappreciated in the EBM literature and is certainly ignored by the evidence hierarchy espoused by EBM. Its lack of inclusion in the PICO/PICOT process also calls into question whether EBM provides a sufficient theory of medicine.

Second, given that the six-step process was meant to accomplish EBM’s goal of standardization, it is necessary to acknowledge that much of the process relies on the subjective – although informed by clinical expertise—judgment of individual clinicians. Consequently, the use of the research literature does not guarantee that all clinicians will assess the literature in exactly the same way or judge the same evidence as best. While referencing the research literature does ensure all clinicians will have access to the same evidence, it does not ensure that the same clinical judgments will be reached for similar patients. The introduction of evidence-based clinical guidelines does go some distance in formalizing and standardizing the decision process. However, EBM does not achieve a higher level of standardization than the previous theory of medicine, as guidelines and clinical references informed by RCTs existed to aid clinical practice in the “old practice” of medicine as well. It is not the case that there is a

---

27 This point will be discussed further in chapter 2.

28 See: Solomon, Making Medical Knowledge. In this book Solomon gives a detailed history of consensus conferences, the guidelines they produce, and their overlap with and relationship to EBM. Solomon successfully demonstrates that clinical practice guidelines have long had a place in medicine.
universal set of guidelines or even relevant guidelines for every clinical encounter. In many ways this raises questions about how different the old and new theories of medicine really are; a question that will be addressed later.

II.11 Dangerous Dogmas & Distrust of Authorities

The second motivation for the development of EBM pertains to the concern expressed by clinicians regarding the dogmatic practices employed in the old theory of medicine. In general, the concern originates from a purported overreliance on clinical experience and clinical expertise as sufficient grounds for medical decision-making. In the EBMWG’s 1992 landmark essay, they contrast the clinical practice of EBM with that of the “former paradigm.” The problematic assumptions of the old paradigm they describe as follows.

1. Unsystematic observations from clinical experience are a valid way of building and maintaining one’s knowledge about patient prognosis, the value of diagnostic tests, and the efficacy of treatment.
2. The study and understanding of basic mechanisms of disease and pathophysiologic principles are a sufficient guide for clinical practice.
3. A combination of thorough traditional medical training and common sense is sufficient to allow one to evaluate new tests and treatments.
4. Content expertise and clinical experience are a sufficient base from which to generate valid guidelines for clinical practice.²⁹

These assumptions present a picture of the old paradigm of medicine as over-reliant on conventional clinical knowledge and basic medical training. Considering that the practice of EBM still acknowledges the necessity of clinical expertise and clinical experience for the practice of medicine,³⁰ the EBMWG’s objections to these assumptions cannot be on the basis

²⁹ Evidence Based Medicine Working Group, “Evidence-Based Medicine,” 2421.
³⁰ The value of clinical experience and clinical expertise within evidence-based medicine is a complex and contentious issue. The literature produced by the EBMWG, its members, and other proponents of EBM is unclear and oftentimes contradictory regarding the appropriate status and function of these sources of knowledge in EBM.
that these sources of knowledge are completely unreliable or irrelevant in medical decision-making. Rather the claim is that when these sources of knowledge are given priority the results are dangerous.

To properly understand this concern, it is useful to consider vignettes from the clinical experiences of a few of EBM’s biggest advocates. In particular, Sir Iain Chalmers and David Sackett, both important figures in the EBM movement, have described moments while treating patients when they experienced serious skepticism of, disappointment in, and failure of the then current medical system (i.e., the theory of medicine in place prior to the advent of EBM) and its educational system.

Sir Iain Chalmers, while working at a refugee camp on the Gaza strip, describes his revelation that the protocol he was taught in medical school for treating measles, which instructed one not to prescribe prophylactic antibiotics, was lethal when treating Palestinian children, who commonly develop rapid bacterial superinfections that can be prevented through the use of prophylactic antibiotics. He recounts:

It made me wonder whether what I had been taught at medical school might have been lethally wrong, at least in the circumstances in which I was working, and precipitated a now incurable “scepticism” about authoritarian therapeutic prescriptions and prescriptions unsupported by trustworthy empirical evidence.31

Sir Iain Chalmers’ story is a prime example of the unease and distrust of the, then, current model of medical education32 experienced by some clinicians. Practitioners began to wonder whether the way medicine had always been taught was the way it should be taught.

---


32 Chalmers attended medical school and practiced primarily in England. His personal recounts are laments over the practice of medicine in the U.K., however, there are similar recounts from clinicians educated in the U.S., Canada, and other Western countries. As such, Chalmers’ sentiments should be understood as generalizable to the
David Sackett, a Canadian clinician, has also described a number of experiences with patients and senior clinicians that brought him to doubt the then current educational model for medicine as well as the sources of expertise clinicians relied on when making treatment decisions.

The most formative and convincing clinical encounter for Sackett occurred during his final year as a medical student. Sackett recalls being charged with the care of “a teenager with ‘infectious hepatitis’ (now called ‘Type-A hepatitis’)” and the subsequent interactions they had:

He presented with severe malaise, an enlarged and tender liver, and a colourful demonstration of deranged bilirubin metabolism… after a few days of total bed rest his spirits and energy returned and he asked me to let him get up and around. In the 1950s, everybody ‘knew’ that such patients, if they were to avoid permanent liver damage, must be kept at bed rest until their enlarged liver receded and their bilirubin and enzymes returned to normal… I wanted to understand (for both of us) how letting him out of bed would exacerbate his pathophysiology. After exhausting several unhelpful texts, I turned to the journals… [and found] “The treatment of acute infectious hepatitis. Controlled studies of the effects of diet, rest, and physical reconditioning on the acute course of the disease and on the incidence of relapses and residual abnormalities.” Reading this paper not only changed my treatment plan for my patient, it forever changed my attitude toward conventional wisdom… [The study had allocated] soldiers who met pre-defined hepatitis criteria at random either to bed rest… or to be up and about as much as the patients wanted… throughout their hospital stay. The time to recovery (as judged by liver function testing) was indistinguishable between the comparison groups, and no recurrent jaundice was observed.

Armed with this evidence, I convinced my supervisors to let me apologize to my patient and let him be up and about as much as he wished. He did, and his clinical course was uneventful.33

Similar to Chalmers, Sackett experienced a failing of clinicians’ “conventional wisdom” during the treatment of his patient. While the consequences of Sackett’s error were far less serious than practice of Western medicine. The same holds true for Sackett, a Canadian clinician. Medical education among Western countries does admit of differences, but the facets of medical education, expertise, and practice causing angst in these vignettes are universal. The similarity of practice among different countries is demonstrated by the fact that they all publish in the same medical journals (e.g., *The New England Journal of Medicine, The Lancet*) and therefore necessarily share the same scientific community, methods, and knowledge.

33 Sackett, “1955 Clinical Trial Report.”
those of Chalmers’, it remains the case that conventional wisdom failed them both. Sackett’s error immediately revealed to him the value of consulting the research literature, convincing him that clinical practice must be guided by current research literature. For both Sackett and Chalmers the advent of EBM was still many years off at the time of these experiences, however both clinicians have retold them in the literature as being pivotal moments for their approach to clinical care and in hatching their desire for a better practice of medicine.

Given the stories of Sackett and Chalmers and the skepticism that resulted, it appears they are making an implicit argument that harkens back to an old critique of appeals to authority in science advanced by Sir Francis Bacon (1561-1626). Bacon held that any uncritical appeal to tradition or one’s predecessors was fallacious because one did not engage in an empirical investigation of nature and did not attempt to establish truth through inductive methods. While this critique of authority may have held in Bacon’s day, it is an unfit critique of contemporary medical practice because the authorities of similar status in medicine today engage in rigorous scientific enterprises and have acquired the necessary expertise to be held in such high esteem. However, EBM does seem to advance a similar objection against appeals to authority. This critique of traditional medicine is evidenced in the vignettes of Sir Iain Chalmers and Dave Sackett, which are employed frequently to tell the story of what motivated the development of EBM. In these vignettes clinical mentors, guidelines, or standard practice are explicitly implicated as the cause of tragic outcomes for patients. In their stories, the authority they referenced or deferred to led them astray, however such instances do not justify a widespread critique and distrust of appealing to such authority.

At this point it is useful to distinguish between two types of authority and how each of these authorities are necessary to the practice of medicine and therefore cannot be deemed
inappropriate or fallacious. The distinction to be made is between someone who is *in* authority and someone who is *an* authority. A person or entity that is *in* authority has the right (because they/it has been granted this right) to enforce the proclamations they make or those they have been put in charge of enforcing. In medicine, examples of being *in* authority include professional and clinical practice guidelines, public health officials and their orders, and hospital or specialist practices. A person or entity that is *an* authority does not necessarily have enforcement powers—although it is not uncommon for *an* authority to be put *in* authority—but is rather considered to have the appropriate training, qualifications, and/or intellectual reputation to be considered an expert in the field and therefore warrants being believed. In medicine, examples of *an* authority include textbooks, institutions like the Cochrane Collaboration, senior clinicians, and specialists.

Medicine’s knowledge base is best described as being the consequence of social epistemology and therefore it is necessary for individuals and entities to exist that are regarded as *an* authority in their particular domain of expertise. Without this type of authority, clinicians would have to subscribe to some version of epistemic individualism, which appears to be the case for earlier versions of EBM, making the enterprise of medicine impossible. Further, to achieve any reliable standards for medicine and its practice, it also has to be the case that there are individuals and entities *in* authority to ensure that all relevant parties are informed of and act in accordance with such standards. Constructing a theory of medicine that prevents appeals to either of these authorities is impractical at best. In an applied science one has to act *as if* one view of that science were true and appeals to authority by clinicians allows them to hold the view of that authority as scientific truth for the purposes of practice. This does not mean intellectual debates cannot be engaged in by experts to develop or revise a particular authoritative view, but this is not the concern of a clinician seeking practical guidance.
From the vignettes provided by Chalmers and Sackett and the EBMWG’s exposition of the “old paradigm’s” assumptions, it is clear that prior to EBM’s inception at least some of its founders and other clinicians had developed a serious skepticism regarding the independent use of clinical experience or clinical expertise as evidence capable of establishing the clinical effectiveness of a treatment. More generally, these clinicians were distrustful of medical authority (i.e., an authority), whether the authority of senior clinicians or the pathophysiological rationale presented in textbooks. Consequently, the status of clinical expertise and clinical experience within evidence-based medicine has been the source of much contention. Efforts to incorporate these forms of knowledge into EBM have been various and for the most part imprecise. Additionally, the aversion to authority in medical practice has also caused complications for EBM because it creates an impractical epistemological standard.

II.III. Expert Consensus

A third and related motivation for the development of EBM was the desire of its proponents to move away from expert consensus, in particular consensus conferences, as the dominant method for settling controversial medical matters. Within the U.S. there has existed more than one type of consensus panel; however, the National Institutes of Health (NIH) Consensus Conferences proved to be the most authoritative and, therefore, will serve as the primary example of expert consensus throughout this project. Each NIH consensus conference included 20-30 expert presentations (with Q&A) for panel members and the public, followed by closed executive sessions during which panel members would craft a consensus statement. The initial consensus statement would then be read to conference attendees followed by comments and discussion. Afterwards the panel would have two hours to make any desired revisions before presenting the
revised consensus statement and their major conclusions at a press conference. In the following month another round of revisions would be made before the final consensus statement was published and distributed.\textsuperscript{34}

From this brief description of what constituted expert consensus (NIH Consensus Conferences were retired in 2013\textsuperscript{35}), there is no mention (beyond the expert presentations) of any type of evidence being consulted, especially not a systematic review of the literature or even copies of the most relevant studies for the panelists to review. This lack of attention to the literature, similar to the lack of attention to the literature individual clinicians like Sackett and Chalmers complained of, was problematic and dangerous in the minds of EBM proponents and, therefore, they deemed expert consensus untrustworthy. Trisha Greenhalgh perfectly captures the essence of how expert consensus panels were regarded by EBM proponents, when she refers to their practice as the “Good Ole Boys Sat Around a Table (GOBSAT) Approach.”\textsuperscript{36} As Miriam Solomon states, “it seems strange to expect 10-20 physician-researchers seated around a table to come to a worthwhile consensus in 48 hours – with no additional research – on a scientifically controversial matter.”\textsuperscript{37} Part of the concern here is over authority, expertise, and a desire to “democratize” medicine, i.e., to make it the case that anyone (assuming proper credentials) can contribute to scientific knowledge, not just a select few experts chosen to represent their field as part of a consensus panel. The other part of the concern is that proper attention is not paid to the medical literature and the “best evidence” it provides.

While the NIH eventually commissioned evidence reports from Evidence-Based Practice Centers funded by the Agency for Healthcare Research and Quality (AHRQ) to be distributed to

\textsuperscript{34} The details of this process are taken from Solomon, \textit{Making Medical Knowledge}, 29-30.
\textsuperscript{35} For further details see: Ibid., 58.
\textsuperscript{37} Solomon, \textit{Making Medical Knowledge},13.
panel members in advance of their meetings, this did not occur until 2001.\(^{38}\) Prior to the commissioning of evidence reports, there was no guarantee that panelists would be presented with the current best evidence and, therefore, the best and most recent medical knowledge wasn’t necessarily considered throughout the process of reaching expert consensus. This neglect of the medical literature once again served as a cause for concern and a motivating factor for the advent of EBM.

\textit{II.IV. Financial Concerns}

The last motivation for bringing about EBM was an acknowledgement of the growing expense of medicine with diminishing returns for patient outcomes. Jeremy Howick describes the financial state of affairs preceding the advent of EBM as follows,

> the cost of healthcare continued (and continues) to rise each year, while improvements in healthcare (measured in life expectancy and infant mortality) have tapered off. Against this background, many thoughtful clinicians began to question the value of the treatments they prescribed.\(^{39}\)

Similarly, in 1991, the year in which EBM was being conceived, E. Haavi Morreim reported that “over the past three decades the total cost of health care has risen exponentially,” so much so “the United States [is spending] more on healthcare than any other nation. And yet … [it] does not necessarily bring us better healthcare.”\(^{40}\) Such concern over the practical value of healthcare dollars has been persistent for decades within the American healthcare system. EBM’s founders had hoped to address the problem by increasing the efficacy and efficiency of the healthcare system through the standardization of the diagnosis and treatment process and acquiring better

\(^{38}\) Solomon, \textit{Making Medical Knowledge}, 50.


“scientific” knowledge of the treatments prescribed through the use of research evidence. However, it is unclear whether the move to EBM has alleviated the situation.\textsuperscript{41}

\section*{II.V Conclusions: Implications for EBM’s Definitions & Development}

The combination of these concerns was enough to motivate the founders of EBM to create a “more scientific” practice of medicine. Evidence-based medicine was the decided answer to this demand, but it remains to be seen whether it was the correct answer and to what extent it actually differs from the previous theory of medicine.\textsuperscript{42}

The initial presentations of EBM in the literature, as well as the concerns motivating its development, have naturally shaped its evolution both in theory and in practice. The primary principles that can be extracted from the origins of EBM that have persisted over the last 23 years are 1) randomized controlled trials (RCTs) are the gold standard research methodology, 2) clinicians should distrust authorities, 3) clinical experience and clinical expertise are not high quality sources of evidence, and 4) standardization of medical practice is an ideal. These four principles have been set forth with varying degrees of rigor, both of strength and detail, throughout EBM’s history and have also been interpreted in a number of different ways, causing much controversy and many complications in the literature.

When assessing EBM, it is fair to hold these principles as the benchmark for the successful practice of EBM because they have been set forth as the implicit goals of the movement along with the general principles of improving patient outcomes and decreasing cost, which are measures of success for any theory of medicine. Even though these principles have not

\textsuperscript{41} The success of EBM’s attempt to lower costs while increasing positive patient outcomes is discussed in further detail in chapter 2.

\textsuperscript{42} The final section of chapter 3 considers whether the current practice of EBM actually achieves any of these goals/motivations.
always been consistently held or enforced to the same degree in the literature, they remain a measure of success for the EBM movement. While some may object to the use of these principles as a measure of EBM, it will be demonstrated (sometimes implicitly, other times explicitly) that these principles are at the root of the different definitions of EBM. However, even though these principles are central to EBM definitions, that does not mean they are well justified or successful at achieving their aims. While there are key features to EBM, the question remains as to how we can define the limits of EBM or exactly what it has come to signify in the literature and in practice.

III. What is Evidence-Based Medicine?

Upon initial contact with the concept “evidence-based medicine,” many find it intuitive. That is, it appears to makes good sense that medicine should, in fact, be based on evidence. On the other hand, more senior clinicians frequently bristle at its mention because they find it somewhat offense: “What do you think we were basing our medical decisions on before? Witchcraft?” Such questions capture the perceived arrogance of the EBM movement and the terms by which it understands itself. However, while EBM may be perceived as intuitive or offensive by some, the real question is, beyond surface impressions, “what is evidence-based medicine?”

In the most basic sense, EBM is concerned with how medicine is understood and practiced. However, it is not a clear and concise theory of medicine. Not only has EBM been a highly contentious movement in medicine, it is recognized, by both its proponents and critics, that it lacks a solid theoretical foundation and first principles. Consequently, as Michael Loughlin puts it, “getting a straight answer to the question ‘what is EBM?’ has proven

---

43 The author has frequently received such questions when discussing this project with senior clinicians at the Texas Medical Center.
surprisingly difficult. Defenders of EBM have frequently ignored criticisms, or responded to them indirectly, by altering their ‘definitions’ of what EBM ‘is and isn’t.’ As such, trying to parse out the theories and definitions attributed to EBM over the last 23 years, both by its founders and other proponents, has proven a great challenge. Additionally, because EBM’s founders never bothered to establish and clarify its first principles or define its cardinal terms, there is no precise construction of EBM that can sufficiently serve as an anchoring point. As mentioned above, the only consistent terms by which we can evaluate what falls under the aegis of EBM are the four principles extracted from its motivations and initial presentation.

It is a primary contention of this project that EBM cannot be conceived of as a singular, unified theory of medicine. Rather, EBM should be understood as a brand name or, like Wittgenstein’s explanation of games in his *Philosophical Investigations*, a cluster concept. That is, there are specific features of EBM—the four principles mentioned above—that are neither necessary nor sufficient on their own to identify a practice or theory as evidence-based. It may appear there is a tension between using these four principles as an evaluative framework for EBM while also conceiving of EBM as a cluster concept. This tension is relieved if one recognizes that EBM has characterized itself as adhering to these four principles and, therefore, is susceptible to being critiqued on these grounds. However, as will be seen, when one looks at the details of EBM in the literature and in practice, many of these principles do not hold, thereby making it accurate to describe what EBM *is* as something like a brand name or cluster concept.

---

44 This is a reference to an article written by David Sackett and colleagues claiming to clarify what EBM is. See: D. Sackett, W.M. Rosenberg, J.A. Gray, R.B. Hayes, and W.S. Richardson, “Evidence-Based Medicine: What It Is and What It Isn’t,” *BMJ* 312 (1996): 71-72.
46 Andrew Miles and Michael Loughlin originally suggested understanding EBM in terms of a “brand name” in a series of papers they’ve published individually and with co-authors.
That is, EBM believes it achieves the preservation of these four principles in all of its instantiations; however, after careful analysis, it can be shown that it does not. At best, it loosely preserves them. This consequence is largely due to the fact that, as Loughlin mentions, many of EBM's defenders have responded to criticism by merely altering definitions or indirectly responding to them, which has served to undercut many of EBM’s principles. As such, there is a vast difference between what EBM says it is and what it actually is.

In the chapters that follow three distinct theories of EBM are described and analyzed: 1) a theory of clinical practice; 2) a theory of clinical decision-making; and 3) a theory for evaluating research evidence of clinical effectiveness (i.e., “rules of evidence”). In each of these theories one (or more) of the principles above is violated or contradicted, lending credence to the claim that it is most apt to understand EBM as a brand name or cluster concept.

To flesh these theories out further, EBM as a theory of clinical practice is the most general theory and the most closely associated with the original definitions of EBM put forth in the literature. Put simply, this theory indicates how various sources of knowledge should be weighted and acted on in clinical practice. It is distinguished from the theory of clinical decision-making advanced in the EBM literature by its inclusion (or consideration) of the patient, the clinician, institutional factors, and clinical factors. That is, EBM’s theory of clinical decision-making focuses on how clinicians should make clinical decisions through consideration of the available evidence of clinical effectiveness with a primary focus on empirical evidence. It should be noted that the rules of evidence are taken for granted and interpreted categorically in the theory of clinical decision-making and therefore is divorced from EBM as a theory for evaluating

---

48 One might think that the accounts identified should more appropriately be considered different parts of one larger theory since they each address a different component of medicine. This claim is rejected because the three accounts outlined employ conflicting assumptions and have developed somewhat in isolation from each other making it difficult to claim that together they comprise a unified theory of medicine.
research evidence of clinical effectiveness. EBM’s theory of clinical decision-making instructs clinicians on how to find and evaluate the best research evidence, again with a focus on empirical evidence. Brief consideration is given to how research evidence can be externalized to a particular patient, but patient desires, etc. are not considered. It is closely tied to the PICO/PICOT process explained above. Lastly, EBM’s rules of evidence constitute a theory of how to evaluate research evidence of clinical effectiveness. The primary focus of this theory is the iconic evidence hierarchy and the epistemological assumptions that support it. It is also different from the other two theories insofar as it has received attention primarily from philosophers of science whereas primarily clinicians have discussed the other two theories in the clinical literature, although a few philosophers have participated.

The ways in which these theories fail to achieve the principles originally set forth by EBM can be broadly understood through the explanation of a few tide changes in the literature. First, when evidence-based medicine was originally presented in the literature it was very much focused on the individual practitioner and the changes one should make in seeking out and critically assessing the best research evidence. The PICO/PICOT process essentially captures this perspective. As was noted in discussion of the PICO/PICOT process above, systematic reviews have become widely available making it unnecessary for clinicians to engage in critical appraisal. This change reduces the clinical decision-making process to formulating a question, locating relevant systematic reviews, and determining how the evidence applies to one’s patient. The clinician is no longer in contact with the original studies and their data. It’s clear from the literature that clinicians could not find the time to employ the PICO/PICOT process in the clinic due to a lack of time, a surplus of studies to wade through, and a lack of statistical skills and knowledge of trial methodology. The development and use of systematic reviews was something
of an ad hoc solution to a significant problem for clinical decision-making and EBM wisely accepted it as a part of EBM. However, there was no justification or attempt to reconcile the use of systematic reviews with the previous conception of EBM and its main principles. In particular, the principle asserting that clinicians should distrust authorities is in direct tension with the use of systematic reviews. That is, systematic reviews replace senior clinicians and textbooks as the authority clinicians consult. The tide turn and tension described here will be further addressed in chapters two and three.

Another tide change in the literature most relevant for EBM’s theory of clinical practice is a shift from focusing on empirically oriented research evidence to its integration with clinical expertise and experience, clinical factors, patient values, and pathophysiological rationale. This shift in the literature has been one of the most contentious because EBM founders simply adjusted the definitions of EBM in response to criticisms that necessary components of clinical practice were missing from EBM theory without providing any justifications for the changes. In addition, by changing the definitions in these ways, questions arose regarding whether clinical experience and clinical expertise could appropriately be considered forms of evidence, knowledge, or clinical warrants within EBM. Either way, it is clear that by being considered an independent form of knowledge/evidence/warrant to be integrated with empirical clinical research, clinical expertise and clinical experience are no longer considered poor quality evidence or, even more problematic, no longer even considered to be types of evidence. In either case, both lead to a violation of EBM’s primary principle requiring clinical experience and clinical expertise to be regarded as poor quality evidence. The issues described here are further fleshed out in chapter two.
Lastly, the principle requiring that RCTs should maintain a gold standard status among all research methodologies has varied greatly in the research literature with proclamations being made that RCT evidence should be considered to the exclusion of all other methodologies and also that EBM cannot justify its cardinal ranking of RCTs. The creation of the evidence hierarchy encapsulated the rules of evidence advanced by EBM. However, there are more than 100 different hierarchies of evidence in circulation making it the case that universal agreement over the ranking schema has not been reached. The most important arguments pertaining to the evidence hierarchy however are those advanced by philosophers of science questioning the epistemological assumptions underpinning the gold standard status of RCTs. The arguments set forth in chapter four will argue that the gold standard status of RCTs is, in fact, unjustified and, therefore, it should not be used as the yard stick by which to measure all other methodologies on the evidence hierarchy. That is, the RCT is not the only methodology capable of achieving the epistemic values of reliability and objectivity nor should they be the only values valued by the evidence hierarchy.

Considering the changes in the literature mentioned above and the foundational principle violated by each, it stands to reason that EBM should be considered a brand name or cluster concept since all of the principles, except standardization, only hold some of the time making them non-necessary conditions. It’s also the case that the principles are not sufficient conditions because any of them taken individually would permit too much variance; that is, clinicians employing vastly different practices could identify themselves as practicing EBM making it non-distinct. Additionally, if one insists on considering the three theories here outlined—1) a theory of clinical practice; 2) a theory of clinical decision-making; and 3) a theory for evaluating research evidence of clinical effectiveness (i.e., “rules of evidence”)—as different parts of a
singular, unified theory, it would make it the case that EBM failed to uphold any of its principles except the ideal of standardization and, therefore, a theory of medicine permitting wildly different practices. The primary contention of this project is that EBM is indeed a cluster concept, however when examined on their individual merits it will also be found that the individual theories of EBM fail on their own terms.
CHAPTER 2
A THEORY OF CLINICAL PRACTICE: DEFINITIONS & CRITIQUE

I. Introduction: Controversies in the Literature

Evidence-based medicine, since its initial articulation in 1992, has been conceived of in many ways. Proponents and critics alike have acknowledged that the concept, “evidence-based medicine” has come to refer to many things – an “educational movement,” “a replacement for traditional medicine,” “a new paradigm for medical practice,” “a method to generate and assess research evidence on the clinical effectiveness of various treatments,” “a subset of evidence-based health care,” “a movement,” “a set of tools and resources for finding and applying current best evidence,” “a vacuous platitude,” and many more.

The primary reason identified in the literature for such wild variations in how EBM is understood, according to critics, is because “getting a straight answer to the question ‘what is EBM?’ has proved surprisingly difficult. Defenders of EBM have frequently ignored criticisms, or responded to them indirectly, by altering their ‘definition’ of what EBM ‘is and isn’t’.”

Proponents of EBM have responded differently to this charge. David Sackett, described as

---

51 Evidence Based Medicine Working Group, “Evidence-Based Medicine,” 2420.
52 Charles, Gafni, and Freeman, “The Evidence-Based Medicine Model,” 597.
“evangelistic [in his] style of leadership” in advocating for the EBM movement, refused to respond to such criticisms stating, “Once again my conclusions came to be given too much credence and my opinion too much weight.” He further claimed that he would “never again lecture, write, or referee anything to do with evidence based clinical practice. [His] energies are now devoted to thinking, teaching, and writing about randomised trials.” This complete disengagement by Sackett is obviously unhelpful for clarifying EBM and has reinforced critics’ judgment that EBM is indeed a “vacuous platitude.”

The most common response offered by proponents for a number of years, however, was to insist that EBM was “evolving” and therefore any theory or definition set forth shouldn’t be taken as definitive. This caused many problems, not because the claim that EBM was evolving was untrue, but because changes in the definitions and theories were never explained or justified. They were simply introduced if as self-evident, confusing many attempts to understand and/or practice EBM. In 2002, R. Brian Haynes finally admitted in print that “EBM has long since evolved beyond its initial (mis)conception, that EBM might replace traditional medicine.” Up until this point, EBM proponents had arduously attempted to tie the roots of their “evolving” theory to the landmark 1992 essay they had published, hoping to maintain some semblance of consistency and to maintain their claim that EBM was in fact a new paradigm. While Haynes finally admitted what critics had been reinforcing for decades, he still gave no explanation or justification for why EBM was “now attempting to augment rather than replace individual clinical expertise and understanding of basic disease mechanisms,” why it had shifted the focus

---

58 J. Daly, Evidence-Based Medicine and the Search for a Science of Clinical Care (University of California Press: Los Angeles, 2005), 75.
60 Ibid., 1283.
62 Ibid., 3.
of critical appraisals from the individual clinician to external pre-appraised sources, or why new authorities (i.e., systematic reviews and their creators) were permitted as a part of EBM – a theory actively discouraging appeals to authority.\textsuperscript{63} While some critics are satisfied with the confessions of EBM proponents like R. Brian Haynes, many still see a need for the genesis of EBM’s theory of clinical practice to be carefully parsed and justified.

II. Genesis of EBM as a Theory of Clinical Practice: Analyzing the Literature

In this chapter select essays from the clinical literature, written by members of the Evidence-Based Medicine Working Group (independently and with co-authors), outlining EBM as a theory of clinical practice will be analyzed. The essays selected for analysis are those most commonly cited by EBM proponents and critics alike and therefore are taken to be representative of EBM theory at the time of their publication.\textsuperscript{64} The essays range in publication dates from 1991 to 2002 and collectively chronicle EBM’s genesis as a theory of clinical practice. The analysis ends with the 2002 articles because they mark a dramatic reconceptualization of EBM as a compliment to clinical practice rather than as a theory of it, which effectively marks the end of one EBM discussion, redirecting debate to issues of evidence, particularly the rules of evidence and the evidence hierarchy, the subject of chapter four.

The analyses undertaken in this chapter achieve two aims: 1) to provide a comprehensive and detailed history of the development of EBM as a theory of clinical practice and 2) to indicate when in the development of this theory tide changes occurred in the literature resulting in the violation of one or more of EBM’s original four principles. The analyses of this chapter will

\textsuperscript{63} For the origins of this sentiment see: I. Chalmers, “Scientific Inquiry and Authoritarianism in Perinatal Care and Education,” \textit{Birth} 10 (1983): 151-166.

\textsuperscript{64} The EBM literature is exceptionally vast and has filled the pages of many journals, diverse in their content, ranging from general medical journals, like \textit{JAMA} and \textit{BMJ}, to specialized journals like \textit{EBM Notebook} and \textit{Cancer Control}. Any omissions of seminal works, while the fault of the author, are not intentional.
pinpoint the limitations, misconceptions, and contradictions of EBM’s theory of clinical practice in an attempt to make clear what principles and characteristics EBM displayed at different stages in its development; lending credence to the claim that EBM is only comprehensible as a cluster concept.

The work undertaken in this chapter is exceptionally important because it is very difficult to explain or discuss EBM clearly without understanding the different stages of its development and its attendant assumptions. Additionally, even though EBM’s theory of clinical practice has in effect outgrown itself, many still use outdated concepts and theories when discussing EBM. Clinical practice also has been shaped by its genesis. The reception and effect of EBM on clinical practice, as well as its difficulties will be discussed in chapter 3.

III. Not So Humble Beginnings

In 1991 the term “evidence-based medicine” was introduced into the clinical literature for the first time. The article, “Evidence-Based Medicine,” served as a simple introduction to what Gordon Guyatt called the “way of the future.” In the article, Guyatt uses the case of an internist treating a 70-year-old man reporting fatigue, suspected of having iron deficiency anemia, to illuminate the contrast he will construct between the “way of the future” and the “way of the past.” In his description of “the way of the past” the internist recalls the advice of an attending physician a few years prior when treating a similar patient. His recommendation had been to order two diagnostic tests. On the basis of this recollection, the internist orders the recommended tests and proceeds to make a diagnosis on the basis of their results. The “way of the future” involves the internist recognizing that she does not know “the diagnostic properties of the tests

---

65 Guyatt, “Evidence-Based Medicine,” A16.
she is considering...and conducts a quick, computerized literature search, using the indexing
terms “iron deficiency anemia” and “sensitivity and specificity.” After finding and appraising a
relevant article she learns that one of the tests she was planning to order would not be useful and
uses the guidelines from the article to interpret the results of the other test. The contrast
constructed here is between relying on the authority of a senior clinician and consulting the
literature to educate oneself and to inform one’s decision. Obvious preference is given to the
latter, especially since the first approach leads to ordering an unhelpful diagnostic test and being
unsure of how to most accurately interpret the other test.

While the essay does characterize the old way of practicing as “[looking] to authority,”
the new way is presented as introducing “additional strategies” and “the inclusion of new
evidence into clinical practice.” While there is no explicit justification as to why new evidence
and new strategies are beneficial, it is indirectly suggested that the new strategies would help
uproot factually incorrect knowledge passed among clinicians leading to the improvement of
patient care. It is also suggested that reliance on traditional authorities is imprudent. At the end of
the article EBM is characterized as requiring the “skills of literature retrieval, critical appraisal,
and information synthesis,” although these terms are never defined.

Overall, the article strikes the reader as mostly benign and merely suggestive of an
increase in literature consultations and appraisals as a part of clinical practice. The strong
rejection of appeals to authority that will later become characteristic of EBM is not prominent in
this article, although it is not entirely absent either. Rules of evidence have not yet been
mentioned and therefore a preference for RCTs has not yet taken root. Slight concern over the

---

66 Ibid., A16.
67 Ibid., A16.
68 Ibid., A16.
trustworthiness of clinical experience and clinical expertise as primary sources of evidence has been introduced but not belabored. There is no mention of standardizing practice. All in all, the 1991 presentation of EBM strikes the reader as a friendly amendment to traditional clinical practice even though reference to “the old way” and “the way of the future” does indicate some type of distinction or proposed shift for practice. What is not present is a grand pronouncement of the introduction of EBM as a Kuhnian paradigm shift.

In 1992 the Evidence-Based Medicine Working Group (EBMWG) published their seminal paper, “Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine.” Like Guyatt’s 1991 essay, the EBMWG begin with a clinical scenario, this time concerning what prognosis a patient who has suffered a grand mal seizure should receive. The old way depicts the clinician informing the patient, at the suggestion of a senior resident and attending physician, that the risk of a reoccurrence is “very high.” On the new paradigm the clinician locates a suitable study and gives the patient specific statistics about the likelihood of reoccurrence over time, the benefits of medication, etc. Consulting the literature is again portrayed as improving the accuracy of patient care and appeals to authority are characterized as leading to vague or misguided recommendations.69

Unlike the 1991 article, this essay fervently proclaims evidence-based medicine as a “new paradigm for medical practice” and offers a definition of EBM as

de-emphasiz[ing] intuition, systematic clinical experience, and pathophysiological rationale as sufficient grounds for clinical decision making and stress[ing] the examination of evidence from clinical research. Evidence-based medicine requires new skills of the physician, including efficient literature searching and the application of formal rules of evidence in evaluating the clinical literature.”70

69 However, it is questionable in this particular case whether the clinician could have obtained similarly precise results by consulting a textbook or other authoritative source.
70 Evidence-Based Medicine Working Group, “Evidence-Based Medicine,” 2420.
Here it appears that the three most important parts of EBM practice are: 1) deemphasizing the components of traditional medicine (i.e., intuition, clinical expertise, and pathophysiological rationale), 2) prioritizing evidence from clinical research, and 3) mastering new skills (i.e., the PICO/PICOT process). Stated briefly, the traditional components of medicine are replaced by evidence from clinical research. More specifically, clinical experience and clinical expertise along with pathophysiological rationale are not to be considered highly trustworthy sources of evidence for clinical decision making. Clinical research evidence appears to be the only information required for clinical decision making. This is made even more evident if one is familiar with the “rules of evidence” referenced in the definition, as they make explicit the exclusion of clinical experience and clinical expertise.

In 1989 David Sackett set forth the “rules of evidence” in an article entitled “Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents,” in which he rejects “a synthesis of the experiences of seasoned clinicians”71 as a reliable foundation for determining recommendations for the clinical management of patients (i.e., clinical experience and expertise are not considered trustworthy sources of evidence). Additionally, he sets forth a ranking of clinical trial methodologies based on the assumption that more tightly controlled trials provide more trustworthy evidence. Consequently, he ranks the epistemic strength of the trial methodologies accordingly: 1) RCTs with low false positive rates, 2) RCTs with high false positive rates, 3) non-randomized, contemporaneous controls, 4) non-randomized, historical controls, and 5) no controls, case-series.72 Here the privileging of RCTs as the gold standard for

---

72 Ibid., 2S-3S.
clinical trials finds its roots in the literature and the sources of evidence for clinical decision-making are narrowed significantly.

The EBMWG also explicitly states that

The new paradigm puts a much lower value on authority. 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. The decreased emphasis on authority does not imply a rejection of what one can learn from colleagues and teachers, whose years of experience have provided them with insight into methods of history taking, physical examination, and diagnostic strategies. This knowledge can never be gained from formal scientific investigation.\(^73\)

While the EBMWG has not rejected the value of authority wholesale, they have relegated traditional sources of authority, like clinical expertise and clinical experience, to a role of establishing basic medical facts (i.e., history taking and physical examination) and strategies for arriving at a diagnosis (e.g., differential diagnosis). Thus it appears that the trustworthiness of authorities is thereby limited to the teaching of skills.

The “new skills of the physician” refer to the PICO/PICOT process outlined in the first chapter. This effectively entrusts treatment decisions to the individual clinician only insofar as one bases their decisions on the evidence provided by clinical research. Privileging clinical research in this manner and insisting that the clinician independently engage in the process of critical appraisal to reach a decision should be understood as advocating for epistemic individualism. That is, EBM rejects the idea that appealing to the authority of experts can “provide justification for claims to know [or]...ground[] rational belief.”\(^74\) This is in direct tension with what the EBMWG call the “traditional practice of medicine,” which relies on collective knowledge and permits reliance on experts in the field (e.g., senior clinicians). It is

---

\(^73\) Evidence-Based Medicine Working Group, “Evidence-Based Medicine,” 2421.

also counterintuitive for most epistemologists working in the medical sciences who have by and large come to accept social epistemology as an accurate reflection of how scientific knowledge communities function, that is, they necessarily rely on the expertise of others because one cannot individually gain the expertise required to verify all knowledge claims oneself.

Adopting such a stance is a dramatic change from traditional medical practice as it is a change in how one verifies knowledge claims (i.e., either individually or collectively) and which sources of evidence one considers to be the most trustworthy and reliable. However, the paradigm of medicine itself has not changed. Clinical research, clinical experience, and clinical expertise all existed and informed medical practice before EBM was created and has continued to do so after its creation. What’s changed is the priority and status each has been given. It is not the case that traditional medicine becomes unintelligible from the perspective of evidence-based medicine, as would be the case if they were in fact two separate paradigms. It merely becomes the case that the different components of clinical practice are valued differently. It is necessary to mention this, not to downplay the radical changes in clinical practice being proposed, but to make clear the limitations of EBM, even if its proponents don’t recognize them.

From this instantiation of EBM it is clear that its characteristic features align with those principles of EBM derived from its initial motivations. By invoking the rules of evidence, they are appealing to the superior status of RCTs for making clinical decisions. They have also made clear that appeals to authority lead to imprecise and potentially harmful results on their account, requiring clinicians to eschew authority and adopt a stance of epistemic individualism. Additionally, clinical experience and clinical expertise are not to be considered sources of evidence at all, but rather fact-finding and/or strategic faculties (i.e., clinical skills). There is no mention of standardization in the essay, however the EBMWG states “physicians whose practice
is based on an understanding of the underlying evidence will provide superior patient care.”75

They are so convinced of this fact they assert it as a primary “assumption” of EBM.

IV. The Beginnings of a Tide Change

In 1995, about three years after the EBMWG’s initial extolling of the virtues of EBM, David Sackett and William Rosenberg published an article entitled, “The Need for Evidence-Based Medicine.” In this essay, a number of shifts away from the original articulation of EBM as a theory for clinical practice occur without explanation or justification and tensions arise between statements made in the article.

The article’s summary claims

As physicians…we have always sought to base our decisions and actions on the best possible evidence. The ascendancy of the randomized trial heralded a fundamental shift in the way that we establish the clinical bases for diagnosis, prognosis, and therapeutics. The ability to track down, critically appraise (for its validity and usefulness), and incorporate this rapidly growing body of evidence into one’s clinical practice has been named ‘evidence-based medicine.’76

From this summary it appears that randomized trials are the only source of evidence EBM is concerned with and also the only basis upon which one can make clinical decisions.77 This directly contrasts with the 1992 definition under which RCTs and other systematized studies (i.e., observational studies and case series) are all considered reliable sources of evidence.

Furthermore, reading the last sentence, it appears that the totality of practicing EBM is

75 Evidence-Based Medicine Working Group, “Evidence-Based Medicine,” 2421.
77 In the years following the publication of this article many understandably came to interpret EBM as advocating for the use of RCTs to the exclusion of all other evidence, despite the fact that such a narrow account of clinical practice would never hold. This is primarily because relevant RCT evidence is not available for all medical diagnoses and treatments and they cannot be performed in some cases due to ethical constraints, costliness, rarity of the condition (i.e., not enough statistical power), and/or an inability to answer a particular clinical question with quantitative data.
performing the PICO/PICOT process. If this is an accurate understanding of the summary, EBM is a very narrow theory of clinical practice.

While Sackett and Rosenberg make such strong claims in their introduction, they later in the same article make seemingly more nuanced claims:

Evidence-based medicine is a short-hand term for five linked ideas: first, that our clinical and other healthcare decisions should be based on best patient- and population-based as well as laboratory-based evidence; second, that the problem determines the nature and source of evidence to be sought, rather than our habits, protocols, or traditions; third, that identifying the best evidence calls for the integration of epidemiological and biostatistical ways of thinking with those derived from pathophysiology and our personal experience…; fourth, that the conclusions of this search and critical appraisal of evidence are worthwhile only if they are translated into actions that affect our patients; and fifth, that we should continuously evaluate our performance in applying these ideas.\(^7^8\)

Taking these points individually, the first point claims that patient, population, and laboratory-based evidence should inform our healthcare decisions. The inclusion of evidence other than RCTs suggests a tension between the statements in the article’s summary and what is claimed to be a primary idea of EBM. Additionally, being aware of the rules of evidence and the EBMWG’s previous denigration of laboratory-based evidence (i.e., pathophysiological evidence), it is hard to understand without any explanation why the category of evidence would be expanded.

While this intuitively appears to be the right approach (i.e., including pathophysiological evidence) and also appears on its face to respond to the mounting criticisms regarding the exclusion of pathophysiological rationale from EBM, the statement is exceptionally vague. While it is inarguable that all three of the evidence types mentioned should be included in the EBM process, as it seems impossible to completely exclude any of them and maintain good practice, there is no indication as to how these different types of evidence should be integrated. Therefore, the extent to which they are actually involved in clinical decision making remains

\(^7^8\) Sackett and Rosenberg, “Need for Evidence-Based,” 620-621.
unclear. The 1992 article understood pathophysiology as allowing “the clinician to better judge whether the results [from clinical research] are applicable to the patient at hand and…as a conceptual and memory aid.”79 If this belief still holds, then it appears that although all three types of evidence support clinical decision making, their status and role in doing so is much different. This would resolve the tension present in the article, however it also raises the question of “evidence for what?” That is, what decisions can different types of evidence inform? For example, RCTs are used to support claims regarding the efficacy of treatments. Lab-based evidence however seems restricted to supporting judgments regarding the applicability of trial results to individuals, which is a significantly different task. Consequently, the status of different types of evidence in EBM is unclear.

The second point appears to be advocating for a contextual approach to determining what type of evidence is necessary for answering a particular question. If this is the case, then Sackett and Rosenberg are here acknowledging limitations on RCT evidence by acknowledging that it may not be appropriate for answering all clinical questions. This again stands in tension with the summary statement for the article.

The third point calls for an integration of epidemiological and biostatistical ways of thinking with those derived from pathophysiological rational and personal experience. The way this idea is stated is somewhat puzzling because it references “ways of thinking,” a very vague term. The suggestion appears to be that in identifying best evidence it must be assessed using epidemiological and biostatistical principles as well as the knowledge we have from pathophysiology and clinical experience. Accordingly, pathophysiology and clinical experience

79 Evidence-Based Medicine Working Group, “Evidence-Based Medicine,” 2421.
adopt an evaluative role in EBM very different from their evidentiary status in traditional medicine.

The fourth and fifth points are intuitive and necessary for any theory of clinical practice to be successful.

Given the statements in the summary of the article and the description of the five main ideas of EBM just presented, it is difficult to determine the exact picture of clinical practice Sackett and Rosenberg have in mind. It appears they are advocating for a practice of medicine where systematic research evidence, especially RCTs, is given priority. Pathophysiology is also understood as providing evidence. However, it is unclear what it would be evidence for. Further, evidence should be selected on a contextual basis and clinical experience and pathophysiological rationale are to be considered evaluative skills. If this picture is correct, it is important to note the consequences of clinical experience and pathophysiological rationale being understood as evaluative skills. If, for example, clinical experience is considered a mode of thinking or problem solving and RCTs are considered a type of evidence, they would necessarily be different in type. Accordingly, clinical experience, as a different component of the clinical decision making process, could override RCT evidence because there would be no way to subordinate it as a lesser quality evidence. This may be what is implied by idea #3 when determining best evidence, but it also seems to give substantial authority to clinical experience, which runs counter to EBM’s anti-authority bent.

Further complications arise when one considers additional statements made by Sackett and Rosenberg. Interestingly, the practice of the PICO/PICOT process becomes diversified in this essay. EBM’s theory of clinical practice, up until this article, was individualistic and emphasized the need for clinicians to engage in the critical appraisal process themselves. The
rationale being that the only way to know truly was to interact directly with the literature oneself. By engaging with the literature clinicians would understand the treatments they were prescribing and their decisions would be based on the most current and reliable evidence available. This perspective was also advocated for to prevent the reliance of clinicians on authorities. However, Sackett and Rosenberg state:

Recent developments and evaluations support the view that three EBM strategies can be successful at keeping us up to date. They consist of learning how to practice evidence-based medicine ourselves, seeking and applying evidence-based medical summaries produced by others, and accepting evidence-based practice protocols, developed by our colleagues and augmented by strategies that help us improve our clinical performance.  

From this statement it is clear that EBM is no longer limited to an individualistic practice. Clinicians can now consult systematic reviews (i.e., reviews of the research literature) produced by others and even evidence-based practice protocols (e.g., clinical guidelines) rather than undergoing the process of critical appraisal themselves. It would seem the acceptance of systematic reviews would constitute the existence of a new authority on which clinicians are meant to rely. So, either EBM is giving up its anti-authority roots or it was never the case that the problem was reliance on authorities but rather that authorities in traditional medicine based their judgments on bad evidence. This would narrow the content of EBM such that it would appear its only contribution to a theory of clinical practice would be emphasizing the use of RCTs. Either way the article neither justifies this shift nor makes any attempt to reconcile it with the EBM theory set forth prior to it. Sackett and Rosenberg’s only explanation appears in the form of an *ad populum* argument in which they state “The performance of systematic reviews of therapy is so logical a step in progress towards *evidence-based health care* that it has become the focus of a rapidly growing international group of clinicians, methodologists, and consumers who have

---

80 Sackett and Rosenberg, “Need for Evidence-Based Medicine,” 621.
formed the Cochrane Collaboration.”\(^8\) Therefore, the justification for including systematic reviews as a part of evidence-based medicine\(^8\) seems to be that everyone values them so much that the Cochrane Collaboration was formed to produce them. While it may be the case that systematic reviews were a logical next step because the limited statistical skills of clinicians, their unfamiliarity with trial methodologies, the sheer volume of the literature, and a lack of available time in the clinic made them necessary, an explanation as to how they are suppose to reconcile with prior EBM theory is entirely absent.

The inclusion of evidence-based clinical protocols, or evidence-based guidelines, generates a similar and perhaps more substantial problem. The content of an evidence-based clinical practice guideline includes information about the clinical testing of different drugs or drug regiments, most likely drawn from systematic reviews, and the attendant benefits and risks. However, they also oftentimes provide information about how a patient should be counseled, what information to discuss, etc. The problem of relying on an authority returns once again, and in this case some of the information is even more subjective insofar as it is not based on systematized clinical research. The argument made on behalf of permitting evidence-based clinical practice guidelines falls back on the value of systematic reviews and RCTs, holding that “when ‘clinical guidelines’ and other practice recommendations are based on this level of evidence [systematic reviews] …they become worth following.”\(^8\) This argument seems to reinforce an interpretation of EBM that isn’t actually anti-authority, but is concerned with the quality of evidence supporting any and all clinical judgments. Accordingly, it appears that Sackett and Rosenberg are expanding the practice of evidence-based medicine beyond its

\(^8\) Ibid., 623.
\(^8\) There is no indication in the article that the use of the term evidence-based healthcare is meant to be interpreted differently from evidence-based medicine.
\(^8\) Sackett and Rosenberg, “Need for Evidence-Based Medicine,” 623.
individualistic roots and permitting clinicians to rely on resources generated by others (i.e., systematic reviews).

To be clear, the argument that can be made on behalf of Sackett and Rosenberg is that, in line with the article’s summary, what really matters, that is, the only essential principle for evidence-based medicine, is the privileging of RCTs. Accordingly, systematic reviews would then rightly be taken to be a high quality source of evidence, even higher than RCTs alone, because they are reviews of all available clinical evidence (i.e., RCTs). Further, evidence-based clinical practice guidelines are also to be considered a high quality source of evidence because they are based on systematic reviews.

If this argument is correct, and the only important principle of evidence-based medicine is the inclusion and privileging of RCTs, it becomes doubtful that EBM is actually a theory of clinical practice because its content does not extend to the clinician-patient interaction nor does it take into consideration the preferences of the patient. Without further clarification of the theory EBM is advocating for it is hard to understand what exactly EBM is advocating for.

V. The Tide Change

In 1996, Sackett, Rosenberg, Gray, Haynes, and Richardson published an article entitled, “Evidence-Based Medicine: What It Is and What It Isn’t.” The primary purpose of this article was “to refine the discussion of what [EBM] is and what it is not.” The article itself is very short and is concerned not only with laying out a new definition of EBM, but also rejecting criticisms alleging that EBM is “old hat,” “impossible to practice,” or “restricted to randomized

---

84 Ibid., 623.
86 This objection will be considered in great detail in chapter 3.
trials and meta-analyses” – some of these criticisms having already surfaced in the analyses above. The reason this essay is singled out as the impetus for a tide change in the literature is because it is the first time EBM explicitly recognizes clinical expertise and patient preferences and values as integral parts of clinical practice. EBM’s purview is no longer restricted to the application of best clinical research evidence.

Sackett et al. newly define EBM as “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.”87 From this definition one can see that an emphasis is now explicitly placed on clinical expertise, the individual patient, and current best evidence. Definitions for key terms are also provided. They state,

By individual clinical expertise we mean the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice. Increased expertise is reflected in many ways, but especially in more effective and efficient diagnosis and in the more thoughtful identification and compassionate use of individual patients’ predicaments, rights, and preferences in making clinical decisions about their care.88

Here, clinical expertise appears to be a catchall for a number of important considerations in the clinical decision-making process. First, it is informed by clinical experience and clinical practice tying it closely to what one might consider clinical wisdom. However, the mention of clinical experience here should be noted because it is the first time since the 1992 definition of EBM that clinical experience has had a place in EBM. However, one can see, given that it has been relegated to informing clinical expertise, that it is no longer considered a source of evidence capable of justifying knowledge claims regarding the efficacy of treatments and the like. Here

88 Ibid., 71.
EBM has succeeded in de-emphasizing its use in clinical decision making, which was intended from their initial statement in 1992.

Sackett et al. also include sensitivity to patient preferences, predicaments, and rights as part of clinical expertise. At this point, it is hard to make a case for clinical expertise being anything other than that which encompasses the skills (all skills!) required of good clinicians. However, it also becomes clear that a large part of clinical expertise is vetting and externalizing evidence. This is obvious in the statement that “Without clinical expertise, practice risks becoming tyrannized by evidence, for even excellent evidence may be inapplicable to or inappropriate for an individual patient.”\(^89\) What is peculiar though is that the externalization (or applicability) of evidence normally employs pathophysiological rationale, making it unclear whether this knowledge source is meant to be a part of clinical expertise as well.

Sackett et al. also clarify the meaning of best available external clinical evidence as “clinically relevant research, often from the basic sciences of medicine, but especially from patient centered clinical research.”\(^90\) Again it is clear that clinical research, that is, systematized research, should be given priority, with RCT evidence continuing to be held as the gold standard. This much has not changed from the previous definition. However, in another essay published in 1996, entitled “Transferring Evidence from Research to Practice: 1. The Role of Clinical Care Research Evidence in Clinical Decisions,” Haynes, Sackett, Gray, Cook, and Guyatt state that

Evidence-based medicine focuses on…systematic studies simply because they represent the most advanced stages of testing to ascertain whether the innovations of basic science work, how well they work, and for whom they work when applied in the clinical setting. Thus, evidence-based medicine is not in competition with basic science; rather it depends on it and builds upon it.\(^91\)

\(^{89}\) Ibid., 71.
\(^{90}\) Ibid., 71.
\(^{91}\) Haynes et al., “Transferring Evidence,” 196-197.
From this discussion of evidence, it becomes clearer that the initial inclusion of evidence from the basic sciences was not meant to suggest that the basic sciences can provide independently valid evidence, but rather that it is foundational to other types of evidence and therefore is necessary for clinical practice, but not as a direct consideration in clinical decision making. At this point, the only source of evidence on which clinicians can base their clinical judgments comes from systematized research. For Sackett et al. to suggest otherwise would be to contradict the statements they’ve made regarding the use and roles of clinical expertise (including clinical experience and patient preferences and values) and evidence from the basic sciences.

In the second 1996 essay discussed, Haynes et al. further clarify the terms conscientious and judicious. They state,

> Conscientious means that evidence is applied consistently to each patient for whom it is relevant. Judicious use calls for the incorporation of clinical expertise that balances risks and benefits of diagnostic tests and alternative treatments for each patient and takes into account his or her unique clinical circumstances, including baseline risk and comorbid conditions, and preferences.\(^{92}\)

By defining “conscientious” in this way the value of standardization in medical practice resurfaces. All similar patients should be treated similarly in accordance with the best evidence. The definition of judicious here merely seems to reinforce the idea that clinical expertise should be used to guide the application of clinical evidence.

Consequently, evidence-based clinical practice requires the inclusion of clinical expertise, patient preferences, and research evidence. However, research evidence is the only form of evidence that can support judgments of clinical effectiveness for a specific treatment. Clinical expertise determines the applicability of evidence to the individual patient and clinical preferences vaguely come into play as an added limitation on treatment decisions. That is, it is

---

\(^{92}\) Ibid., 196.
acknowledged that “clinical expertise and patient preferences may override the other components of the model for a given decision.” Accordingly, a more comprehensive theory of clinical practice seems to be emerging.

While clinical expertise and patient preferences have re-entered the EBM picture of clinical practice and the details of the theory have shifted, it remains the case that the theory constructed still adheres to the original principles outlined for EBM. First, standardization has been incorporated through the caveat of best evidence being used conscientiously; second, defining clinical expertise and clinical experience as skills for applying evidence necessarily indicates that they are no longer considered epistemically trustworthy sources of evidence; third, RCTs continue to be the focus of EBM as its gold standard trial methodology; and fourth, traditional authorities have been eliminated from evidence-based clinical practice, however it appears that new authorities have been imported through the acceptance of systematic reviews and evidence-based guidelines.

In 2000, a new definition of evidence-based clinical practice is set forth in Sackett’s textbook, Evidence Based Medicine: How to Practice and Teach EBM, 2nd edition. In this text he states, “evidence-based medicine (EBM) is the integration of best research evidence with clinical expertise and patient values.” This definition is not a dramatic change from the 1996 definition, however it does recognize patient values as a separate component of the decision-making process rather than a component of clinical expertise. This is an intuitive move because the inclusion of patient preferences in the definition of clinical expertise, as was the case in the 1996 definition, implies that the clinician is ultimately in control of managing the patient and one’s preferences, which denies the autonomy of the patient and seems to suggest a paternalist model of medicine.

---

93 Ibid., 197.
one which contemporary medicine and bioethics has condemned and departed from. Beyond this change the particulars of evidence-based clinical practice appear unchanged.

VI. Confessions and Concessions of EBM Proponents

Two important essays published in 2002 dramatically changed the contours of EBM. After two decades of criticism and claims that EBM was an impossible theory of clinical practice too narrow in its definition of evidence, too cumbersome in practice, and too dismissive of the clinical context, patient preferences, and clinical expertise, EBM proponents made some concessions.

R. Brian Haynes appears to be the driving force behind such concessions as he (independently and with co-authors) published the articles that recognize some longstanding problems with EBM theory and dramatically reinterpret it. The most striking admission, penned by Haynes in his article, “Clinical Expertise in the Era of Evidence-Based Medicine and Patient Choice,” was that “EBM [had] long since evolved beyond its initial (mis)conception, that EBM might replace traditional medicine. [Accordingly,] EBM is now attempting to augment rather than replace individual clinical experience and understanding of basic disease mechanisms.”

This acknowledgment stands in stark contrast to the EBMWG’s initial claim that evidence-based medicine was a new paradigm of medicine. Additionally, by giving up the contention that EBM constitutes a unique paradigm of medicine clinical experience and basic disease mechanisms no longer need to be subsumed under clinical expertise or misleading packaged into research evidence or some how converted into a clinical skill. Most importantly, as will be shown, they are once again considered sources of evidence for judging clinical effectiveness (at least in some cases).

---

95 Haynes, “What Kind of Evidence.”
cases). Overall, the new theory of EBM advanced is much more coherent, with its terms well
defined and the roles of the different components of clinical practice clearly outlined.

To further emphasize the new role of evidence-based medicine in clinical practice – as it is no longer a theory of clinical practice – Haynes puts forth a “technical definition” of EBM stating that “evidence from health care research is a modern, never-before-available complement to traditional medicine. Perhaps a better name [for EBM] would be “certain-type-of-high-quality-and-clinically-relevant-evidence-from-health-care-research-in-support-of-health-care-decision-making.”…an accurate but mind-numbing descriptor.” Haynes also advances a “pragmatic definition of EBM” as “a set of tools and resources for finding and applying current best evidence from research for the care of individual patients.” Here the tools and resources he is referencing refer to critical appraisal, systematic reviews, evidence-based clinical practice guidelines, etc. From these two definitions it is clear that EBM is now only concerned with producing, appraising, applying, and integrating high quality research evidence into clinical practice. Limited in this way, it appears difficult to object to EBM. However, Haynes admits that the fundamental assumption declared in 1992 “that practitioners whose practice is based on an understanding of evidence from applied health care research will provide superior patient care compared with practitioners who rely on an understanding of basic mechanisms and their own clinical experience” can not be substantiated. He admits, “So far, no convincing direct evidence exists that shows that this assumption is correct.”

Haynes, Devereaux, and Guyatt, in their 2002 article, “Clinical Expertise in the Era of Evidence-Based Medicine and Patient Choice,” describe the contrast between the original EBM theory and its newer articulation as follows.

96 Ibid.
97 Ibid.
98 Ibid.
Initially, evidence-based medicine focused mainly on determining the best research evidence relevant to a clinical problem or decision and applying that evidence to *resolve* the issue. This early formulation de-emphasised traditional determinants of clinical decisions, including physiological rationale and individual clinical experience. Subsequent versions of evidence-based decision making have emphasized that research evidence alone is not an adequate guide to action.\(^99\)

This contrast is important because it admits that EBM once saw best research evidence as sufficient for “resolving” clinical issues. That is, the traditional components of clinical practice and clinical decision making were viewed as unnecessary. Given the habit of some EBM proponents of citing the initial 1992 article to support claims that EBM acknowledges these traditional components, it is important to recognize this fact.

Under the new *model* of evidence-based clinical decision making, Haynes et al. base decisions “on patients’ circumstances, patients’ preferences and actions, and best research evidence, with a central role for clinical expertise to integrate these concepts.”\(^{100}\) By referring to this as a model it does not purport to include all components of clinical practice as a part of EBM. They state, it “is prescriptive rather than descriptive. That is, it is a guide for thinking about how decisions should be made rather than a schema for how they are made.”\(^{101}\)

There are many other important advances to recognize in this model, For example, this model is the first to explicitly recognize the clinical circumstances of patients. Haynes et al. state “Patients’ clinical state, the clinical setting, and the clinical circumstances they find themselves in when they seek medical attention are key, and often dominant, factors in clinical decisions.”\(^{102}\)

In 1996, the introduction of the term “judicious” to EBM’s definition did identify a need to pay attention to clinical circumstances, however, it was considered to be part of clinical expertise and

---

100 Ibid., 36.
101 Ibid., 36.
102 Ibid., 37.
was not recognized as a separate component as it is in the new theory. This is an important part of the clinical decision making process not only because it provides basic factual information about the clinical case one is presented with, but also because it often limits the diagnostic and treatment options available. While high quality research evidence may be an important factor for making treatment decisions and the like, one of the main reasons it cannot resolve clinical questions independently is that the patient populations used in clinical research normally do not identically match the individual patient being treated. That is, there may be differences in age, gender, ethnicity, race, co-morbidities, etc. that can effect the efficacy of the treatment. Admittedly, “personalizing” the evidence to fit a specific patient’s circumstances is a key area for development in evidence-based medicine,“103 and independently recognizing the importance of patients’ clinical circumstances is a first step.

Haynes and Haynes et al. are very careful to define the terms used to illuminate their model of clinical practice. Some terms carry the same meaning as in previous definitions, but some are importantly different. Clinical expertise, for example, plays the same role as it has in previous definitions: it “includes the general basic skills of clinical practice as well as the experience of the individual practitioner. Clinical expertise must encompass and balance the patient’s clinical state and circumstances, relevant research evidence, and the patient’s preferences and actions if a successful and satisfying result is to occur.”104 However, what is unique about the new model is that clinical experience is not only a part of clinical expertise – useful in balancing the components of clinical practice – it also appears to be allowed an evidentiary role in determining clinical effectiveness. Haynes states “EBM claims that experts are more fallible in their recommendations (of what works and what doesn’t work in caring for

---

103 Ibid., 37.
104 Ibid., 37.
patients) than evidence derived from sound systematic observation (i.e., health care research).”

From this statement it is clear that clinical experience is not considered highly reliable evidence for clinical effectiveness, however it allows for it to be recognized as a source of evidence albeit an imperfect one. This has not been acknowledged in any of the previous articles and is a serious indication that EBM really does regard itself as a compliment to traditional medicine at this point in time.

A seemingly contradictory statement to the claim made above is Haynes’ claim that “evidence is narrowly defined as having to do with systematic observations from certain types of research.” This would seem to preclude clinical experience as a type of evidence. However, Haynes et al. further explain that “research evidence” is to be distinguished from other forms of information that have always been part of clinical decisions, such as the patient’s history, physical findings, diagnostic tests, circumstances, and stated preferences. Research evidence includes systematic observations from the laboratory, preliminary pathophysiological studies in humans, and more advanced applied clinical research, such as randomized controlled trials with outcomes that are immediately important to patients.

Given this qualification of research evidence, clinical experience does fall outside its purview, however it remains the case that it has “always been part of clinical decisions” and therefore should remain a source of evidence under this theory.

Haynes also provides additional information as to what types of studies are considered a part of clinical research beyond RCTs, acknowledging “different methods, observational or experimental, are needed for exploring different questions.” He states, “it is important to recognize that experimental designs were added to observational designs, not substituted for

---

105 Haynes, “What Kind of Evidence.”
106 Ibid.
them.”109 Furthermore, “methodologies from other scientific disciplines have been added…nonexperimental and qualitative research methods have been adopted…thus, the research methods of medical science are pluralistic and expanding.”110 He also notes that “EBM does not clearly address the role of basic science in medical discovery, except to indicate that, in most circumstances of relevance to individual patient care, basic science alone does not provide valid and practical guidance. There are some exceptions.”111 Given these additions to what can be considered evidence and what methodologies are acceptable, EBM is diversified from its typical emphasis on RCTs.

Haynes also addresses the tasks of EBM users (the PICO/PICOT process), acknowledging that an insistence on individual clinicians having direct contact with the literature is “hopeless” and advocating for clinicians to “[find] the right pre-assessed research evidence, [judge] whether it applies to the health problem at hand, and then [work] the evidence into the decision that must be made.”112 While this revelation was made earlier in the genesis of EBM, Haynes’ recognition of the original PICO/PICOT process as hopeless very much indicates that the new model of EBM has moved beyond its roots to a more practical and pragmatic theory.

The new model of EBM proposed by Haynes and Haynes et al. has unsurprisingly abandoned a few of EBM’s foundational principles. First, it is clear from its acceptance of systematic reviews and, more importantly, clinical experience as evidence for clinical effective that EBM is no longer against appeals to authority in clinical practice. Second, clinical experience and clinical expertise while certainly not considered high quality clinical evidence, have been recognized as integral components for clinical practice and therefore now enjoy an

---

109 Ibid.
110 Ibid.
111 Ibid.
112 Ibid.
elevated status. Third, RCTs continue to be the gold standard of research evidence, but serious
effort has been made to recognize that there are not only other reliable types of systematic
evidence, but also that RCTs may not be useful for all clinical questions. Last, this EBM model
does not talk directly about the standardization of medical practice, but it does insist on a causal
link between EBM and increased quality of care, even if one cannot be definitively proven.
EBM’s softening in its adherence to some of these principles and the rejection of others
demonstrates that EBM is now something different from its original theory. However, both are
still considered “evidence-based medicine.” This may indicate why Haynes now considers EBM
a “generic term.”

VII. Conclusion

After analyzing the genesis of EBM as a theory of clinical practice, it is obvious that it has not
consistently fulfilled EBM’s original principles, its definitions and contours have changed
consistently, and, ultimately, it has abandoned its claim of being a theory of clinical practice. The
three primary developments that occurred over the course of its genesis are: 1) abandoning the
individualistic nature of EBM such that clinicians can rely on pre-appraised sources and other
authorities so long as they are informed by high quality evidence; 2) moving from a theory where
clinical research evidence was sufficient for answering clinical questions to a theory that
acknowledged, even if in a strained way, the necessity of clinical expertise, pathophysiological
rationale, and the preferences and clinical circumstances of the patient; and 3) an
acknowledgment that EBM is not and cannot be a theory of clinical practice and therefore is
better understood as a compliment to traditional clinical practice.

113 Ibid.
The analyses undertaken in this chapter end with the contributions of EBMWG members in 2002. The reason for this finishing point is that after Haynes et al.’s admissions and the establishment of EBM as a compliment to clinical practice concerned with the evaluation and integration of best research evidence, the important questions to ask about EBM changed along with the relevant literature. The important questions to ask about EBM now pertain to the rules of evidence and how they are justified, as well as how well clinical research evidence can be externalized to individual patients. These concerns are the focus of chapter four.

Before moving on to an analysis of EBM as a theory for evaluating research evidence of clinical effectiveness, it is necessary to consider how EBM’s theory of clinical practice has been infused into actual practice, its success (or failure) in the clinic, and, most relevant for the analyses just undertaken, the reasons why its earlier articulations were unsuccessful – i.e., given that there are no explanations in the literature justifying the tide turns witnessed in the genesis of EBM, it is productive to see the practical failures that motivated these shifts.
I. Introduction

In chapter two the genesis of EBM’s development as a theory of clinical practice was carefully laid out with particular attention paid to the three distinct theories\textsuperscript{115} espoused by EBM proponents over time. The first theory (1992) defined EBM in opposition to traditional medicine, deemphasizing traditional evidentiary sources (i.e., clinical expertise, clinical experience and pathophysiological rationale) and stressing the importance of clinical research evidence for clinical decision making and the use of the PICO/PICOT process by individual clinicians for assessing evidence. The 1996 definition advocated for the integration of clinical expertise and best research evidence when making clinical decisions and also permitted the use of pre-appraised sources of evidence (e.g., systematic reviews), effectively freeing clinicians from having to conduct critical appraisals of the clinical research literature. In 2002, Haynes radically redefined EBM as a \textit{compliment} to clinical practice focused on \textit{integrating} best research evidence into clinical practice. As such, the best research evidence is seen as an additional

\textsuperscript{114} The reader should be reminded that constructing a universally agreed upon interpretation of EBM for analysis and criticism is an impossible project due to the fact that EBM is a cluster concept admitting of several theories, each of which has undergone its own unique genesis in the literature. Since EBM’s initial introduction in 1992 its formal definition in the literature has changed three times (see: EBMWG 1992; Sackett et al.1997; and Sackett 2000, see chapter 2 for complete definitions and an analysis of their development) and a substantial amount of rhetoric regarding what constitutes best evidence and evidence-based medicine more generally has pervaded practice and the medical and philosophical literatures making the interpretations of EBM among different groups of practitioners, proponents, and critics disparate and diverse. As such, it is an inescapable truth that the interpretation of EBM theory provided in this chapter may conflict with the interpretations of others. Consequently, one cannot claim that the particular interpretation of EBM theory employed here is a strawman. Furthermore, given the ample citations provided from the clinical literature (including authors that are proponents and critics), the philosophical literature, medical textbooks, and even some research studies, it is clear that the interpretation constructed here is at least one popular interpretation of EBM theory. For more detailed comments about this please see chapter 2.

\textsuperscript{115} See: Evidence Based Medicine Working Group, “Evidence-Based Medicine,” 2420-5; Sackett et al., “What It Is and What It Isn’t,” 71-72; Haynes, “What Kind of Evidence.” Also, please see chapter 2 for detailed descriptions and analysis.
component of clinical practice to be considered along side other components from traditional medicine (i.e., clinical expertise, pathophysiological rationale, patient values, and clinical circumstances). This allowed for a diversification in the evidentiary sources clinicians were permitted to seek out in support of clinical decisions.

As was mentioned previously, many of the changes to the definitions of EBM occurred in the literature without justification or detailed explanation as to the motivations for doing so. In this chapter the rationale behind these changes will be established through an examination of how evidence-based medicine, as a theory of clinical practice, has been implemented in the clinic. That is, by examining how actual clinicians have attempted to practice EBM and the practical and conceptual constraints that have hindered them.

To construct actual evidence-based practice, the use of a current EBM textbook – *Essential Evidence-Based Medicine*, second edition, by Dan Mayer – and some preliminary clinical studies of the clinical behavior of medical residents, fellows, students, and practitioners will be drawn upon. While one might expect that since the most recent definition of EBM was set forth in 2002, EBM textbooks and clinical practice would have conformed to its principles by now (i.e., 13 years later). However, practitioners of EBM still tend to vary in their practice, especially regarding the extent to which they independently engage in the PICO/PICOT process and which evidentiary sources they consult. EBM textbooks still teach the PICO/PICOT process to clinicians in addition to addressing the construction and use of pre-appraised sources. Accordingly, the taxonomy of practitioners provided by textbooks includes practitioners that roughly align with the three distinct definitions of EBM (i.e., 1992, 1996, 2002) set forth in the literature. By critiquing the clinical practice of the clinicians in each of the categories set forth by
Mayer’s taxonomy one is able to determine the practical and conceptual constraints that motivated and justified changes to the dominant conception of evidence-based clinical practice.

It will be argued that the 1992 articulation of evidence-based clinical practice was overly severe in its dismissal of appeals to authority and also required an overly narrow focus on critically appraised research evidence as sufficient for informing clinical decisions. Accordingly, these factors are what motivated EBM proponents in 1996 to advocate for the use of pre-appraised sources and clinical expertise when making clinical decisions. However, similar concerns continue to plague the essays set forth in 1996 because pre-appraised sources still require some measure of critical appraisal from the clinician and the evidence gathering behaviors of clinicians demonstrate that they need clinically relevant and easily accessible evidentiary sources, which most pre-appraised sources are not. It is further argued that the recognition that EBM’s 1996 articulation was insufficient as a theory of clinical practice and more appropriately suited for infusing the research literature into clinical practice (i.e., into all evidentiary sources) prompted EBM to redefine itself such that it took on a complimentary role to clinical practice in 2002 allowing it to focus on the dissemination of research evidence.

To gain an indication of whether EBM has been successful in achieving its primary goals, the conclusion of this chapter will consider the degree to which each goal has been realized by current EBM infused medical practice. The purported goals of EBM include: 1) minimizing the inverse relationship between healthcare costs and patient outcomes; 2) eliminating clinicians’ deference to an authority; 3) discouraging the use of clinical experience and clinical expertise; 4) standardizing medical practice; 5) improving patient outcomes; and 6) better integrating clinically relevant research into medical practice. Ultimately, it will be found that either due to a
lack of evidence or the constraints of clinical practice, the only goal to be realized thus far is a *better* integration of clinically relevant research into medical practice.

II. Taxonomy of Practitioners

After surveying the EBM literature relevant to the teaching and practice of EBM in medical schools and hospitals, one is able to discern that there is at least a preliminary recognition of the fact that the practice of EBM\(^{116}\) amongst clinicians, fellows, and residents is not uniform.\(^{117}\) In fact, there are common distinctions drawn between different groups of practitioners based on their relationship to (i.e., interaction with) and understanding of the evidence base, which roughly fall into three separate categories captured by the taxonomy of practitioners put forth by Dan Mayer in his second edition textbook, *Essential Evidence-Based Medicine*. Mayer’s taxonomy includes the categories of doers, users, and replicators.\(^{118}\) The doer type most closely mimics the practice of EBM as it was conceptualized in 1992, what will be referred to as “textbook EBM,” that is, clinicians are themselves engaged in the critical appraisal process, have direct contact with the published results of studies, and prioritize the use of RCTs. The user type practices a version of EBM that adheres most closely to its 1996 conceptualization, that is, they access pre-appraised sources and do not participate in the full process of critical appraisal. Lastly, the replicator type practices EBM in line with its reconceptualization in 2002, accessing

\(^{116}\) It is the argument of this chapter that EBM is in fact impossible to practice. Consequently, the use of the term, “practice of EBM” is meant to indicate attempts to practice EBM.

\(^{117}\) The lack of uniformity I am referring to applies to the group of clinicians, fellows, and residents as a whole. Within the individual groups there is a lack of uniformity and practice. That is, the differences I am pointing to in how one practices do not correlate to differences in education and/or experience but rather to a difference in how they understand and interpret EBM.

\(^{118}\) Depending on the sources one consults, different authors use different terminology to refer to the largely similar distinctions they make among the different ways EBM is practiced. Some authors only distinguish between two groups, not even acknowledging “replicators” as practicing EBM. The terms “doers,” “users,” and “replicators” have been chosen for use because the terms intuitively illustrate the relationship the healthcare professional has with the evidence base and they also lend themselves illustratively to how much authority the healthcare professional assigns to the evidence base.
the most readily available evidentiary sources, but not necessarily regularly consulting or assessing the research literature or pre-appraised sources. The primary evidentiary sources used by replicators (e.g., other clinicians, pocket guides, and textbooks) are those sources whose use in traditional medicine concerned the founders of EBM. However, how well replicators achieve evidence-based practice (i.e., that informed by clinical research) will depend on whether the evidentiary sources they consult draw on clinical research (e.g., a textbook drawing on the best clinical research can be considered good quality evidence). It should be noted that the distinctions Mayer makes between practitioners are somewhat vague and are not well defended, however, they function as a useful heuristic for making distinctions among the various ways EBM has been instituted in practice and they align with three primary conceptualizations of EBM advanced EBM proponents.

In the following sections a brief description of each of Mayer’s categories will be provided and the attendant issues for each category will be discussed. By illustrating the different types of practitioners present in the clinic and the conceptual and practical barriers that arise from their practice, arguments can be made for why it was necessary for EBM’s theory of clinical practice to evolve in the way it has, arguments never explicitly made by EBM proponents in the literature.

II.I “Doers” & Textbook EBM (1992)

For Mayer, there are three types of practitioners of EBM: “doers,” “users,” and “replicators.”119 “Doers” can be taken to most closely approximate the practice of EBM as it was set forth in

---

119 It should be noted that Mayer never explicitly defines these three categories in any detail. Rather he resorts to describing what the practice of each type of practitioner would look like. As such, it is the aim of this chapter to further flesh out these categories and what Mayer likely means as well as using empirical studies to support the definitions constructed in this chapter.
1992, i.e., what will be referred to as “textbook EBM.” Doers have direct contact with the scientific literature and take a critical approach to it, engaging in the entirety of the PICO/PICOT process. In Mayer’s words,

Most practitioners have to keep up by regularly reading relevant scientific journals and need to decide whether to accept what they read. This requires having a critical approach to the science presented in the literature; a process called “doing” EBM and the activity is done by “doers.”

Given Mayer’s description, it is clear that he is assuming that doers of EBM have the requisite knowledge and skills to effectively search the literature base, critically appraise those evidentiary sources that are prima facie relevant to the clinical question at issue, and ultimately can apply the knowledge gained through the process of critical appraisal to the individual patient.

On its face, textbook EBM appears to be a simple method easily translatable into practice. As Gordon Guyatt and colleagues understand medicine: “High quality health care implies practice that is consistent with best evidence. An intuitively appealing way to achieve such evidence based practice is to train clinicians who can independently find, appraise, and apply best evidence.”

The point here appears simple: good health care requires the best evidence; to get the best evidence you train clinicians who can find, appraise, and apply it. However, in reality, teaching clinicians the appropriate practice of textbook EBM and having that practice realized is difficult. As Brendan Reilly confirms, “many of us teach EBM

---

120 Mayer, Essential Evidence-Based Medicine, 2nd, 11.
121 Mayer’s assumptions are justified here only in that he later claims that the purpose of his textbook is to turn readers into doers. The content of the textbook supports the development of all these abilities. However, if using this taxonomy to identify clinicians who are practicing as doers, given the brief definition Mayer provides there is nothing to show clinicians are actually able to implement the knowledge they gain through critical appraisal in the clinical context with an individual patient. The application of population level research to individual patients is widely acknowledged as difficult and sometimes impossible depending on a patient’s physiology, co-morbidities, etc. It is also doubtful that clinicians can actually sufficiently develop the skills to the extent and degree required to conduct nuanced critical appraisals.
122 Guyatt 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.
(integrating best evidence with clinical expertise and patient values) knowing that it is nearly impossible to practise it in everyday clinical care”\textsuperscript{123,124}

Looking to the empirical literature, it appears there is a small contingent of clinicians who have a sophisticated understanding of EBM and are able to realize something close to what has been constructed as textbook EBM in practice. In a survey of the residents of two U.S. pediatric residency programs,\textsuperscript{125} Timmermans and Berg labeled these kinds of practitioners “researchers” in contrast to their less sophisticated peers,\textsuperscript{126} the “librarians.” Under the taxonomy of practitioners given by Mayer, researcher residents would correlate most closely to “doers” and librarians would correlate most closely to “users.” The researcher residents were a small minority and, according to Timmermans and Berg,

Took the core of EBM to mean that the physician acted more as a \textbf{researcher} who actively evaluates and interprets the literature. Residents who profess to be more familiar with EBM specified that merely checking published literature is insufficient but that EBM implies a critical assessment of available evidence in a meta-analysis…For researcher residents, the persuasive strength of recommendations does not depend on where findings are published…Researchers use protocols and review articles as an intermediary step to more specialized evidence…Researchers did not look in the scientific literature for pragmatic guidance to treat the patient at hand, but for a variety of factors to take into consideration during decision making.\textsuperscript{127}

\textsuperscript{124} There still is not good evidence regarding the best practices for teaching EBM to medical students, residents, and clinicians. Some studies suggest that teaching EBM in a traditional classroom setting is ineffective and has better outcomes when integrated into the clinical learning environment. For discussion see: A. Coomarasamy and K.S. Khan, “What is the Evidence that Postgraduate Teaching in Evidence Based Medicine Changes Anything?: A Systematic Review,” \textit{British Medical Journal} 329 (2004): 1017-1021.
\textsuperscript{125} To give further background for this study, Timmermans and Berg describe their sample as based on “in-depth interviews with seventeen pediatric residents of two U.S. pediatric residency programs. Both programs were part of large, urban hospitals affiliated with academic institutions. As residents, our respondents had finished four years of medical school were at different stages of three years of rotations in different clinical pediatric specialties. Most of the respondents were in their mid-20s and white (three Asian respondents). The gender distribution was nine male and eight female residents. Eight respondents were in their first year of residency, two in the second year, five in the third year, and two chief residents were in their fourth year. Their rotations at the time of the interview varied from the newborn unit, endocrinology, hematology-oncology, pediatric intensive care, pediatric surgery, to the emergency department” (S. Timmermans and M. Berg, \textit{The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care}. (Philadelphia: Temple University Press, 2003),144).
\textsuperscript{126} Unsophisticated regarding the practice of EBM only.
\textsuperscript{127} Timmermans and Berg, \textit{The Gold Standard}, 149-150 (bold face mine).
From this description it is clear that something central and distinctive about practicing textbook EBM is undertaking a critical appraisal of the literature. In undertaking a critical appraisal the clinician has an active interaction with the evidence base (i.e., the literature) and derives the power of the evidence, neither from the prestige of the author nor from the ranking of the journal, but from the force of the evidence once it has been vetted through the process of one’s own critical appraisal. Further, because the clinician is the one conducting the critical appraisal, one does not have to trust another’s expertise to carry out the process. Researchers also do not access the literature just for the purpose of gaining practical guidance, but rather to help shape the decision-making process such that the most important considerations are taken into account or for purely unapplied purposes. It is clear that researchers understand EBM as requiring a deep analysis and engagement with the literature, ruling out superficial searches and the mere consultation of published texts.

One can see from this description of doers/researchers that the received understanding of practicing textbook EBM is meant to prevent appeals to an authority, which, as it has already been argued, is an impossible way to practice medicine. By refusing to trust the expertise of authors creating systematic reviews and the journals that publish them, doers/researchers are essentially understanding the practice of medicine as one where one must always verify claims for oneself and can never simply appeal to or rely on other experts or communal knowledge in the field; what has previously been referred to as epistemic individualism.

While it is built in to the definition of doers that they have the requisite skills to practice EBM, most clinicians do not actually possess the necessary skills of critical appraisal nor do they have the time to conduct them, necessitating the existence of authoritative evidentiary sources for clinicians to appeal to. This is most likely the dominant reason why, contra the approach in 1992,
there was a shift in the literature in 1996 to include pre-appraised evidence. Recall that R. Brian Haynes even admits that textbook EBM was “hopeless.”

To emphasize the practical impossibility of doers, and consequently why the 1992 articulation of EBM could not be sustained, it is necessary to consider the clinical context in further detail. Two primary, practical difficulties for implementing textbook EBM include its time consuming nature and the difficulty of inculcating clinicians with the sophisticated skills required for its successful practice. However, it should be mentioned that most of these constraints are symptomatic of EBM requiring clinicians to engage in critical appraisal when they are not experts in critical appraisal and cannot reasonably be expected to conduct a critical appraisal in the same manner or of the same quality as one undertaken by a professional (expert) clinical appraiser (e.g., someone working for the Cochrane Collaboration). The following sections will outline the practical challenges clinicians face – time constraints and a lack of the requisite skills – and will draw on some statistics from clinical practice to illuminate the severity of each challenge.

II.I.I. Time Constraints

The factor overwhelmingly cited to explain the difficulty of practicing textbook EBM is the constraint of time. In a study of EBM-trained medical residents, McCord and colleagues

---

129 McCord et al.’s study was “conducted at five hospital-based family medicine residency programs affiliated with the Northeastern Ohio Universities College of Medicine in Rootstown, Ohio. The office of research of the department of family medicine designed and supervised the study. The office consists of the department chair, six faculty physician research directors from the residency sites, a community based physician, a research coordinator/statistician, and an administrator. Study participants were all 25 third-year residents practicing family medicine at the five participating residency sites.” (G. McCord et al., “Answering Questions At the Point of Care: Do Residents Practice EBM or Manage Information Sources?” Academic Medicine. 82 (2007): 298.)
130 According to McCord et al., the residents involved in the empirical study “in their second year of study…had completed a required EBM curriculum during which EBM principles were informally taught and discussed during the delivery of patient care, seminars, journal clubs, self-study projects, and research or scholarly projects. Toward
found that “rather than practicing [textbook] EBM, residents operated more as information managers [i.e., users and replicators] within the constraints of time limitations and job responsibilities.”\textsuperscript{131} Assuming that the training residents received enabled them to be EBM doers (although this has been shown to be unlikely), it appears that even those who had the skills and abilities to practice EBM were constrained from doing so because of the overwhelming demands on their time. Further, in Singh and Oswald’s introduction to evidence-based practice, they cite an article by Davidoff and colleagues, which gives staggering statistics regarding the time it would take for clinicians to stay current with the research literature. They state:

Busy clinicians cannot keep abreast of the changing architecture of possible interventions when one compares the time required for reading pertinent research to the time they actually have available for the task. For example, physicians would need to read 19 journal articles a day, 365 days per year to keep abreast of the current literature in general medicine.\textsuperscript{132}

This statistic is from an article published in 1995. The number of trials conducted has proliferated each year since the early 1990s and there is no indication that the numbers are in decline or will be soon. A more recent example demonstrates that there is still an overwhelming and unmanageable wealth of information relevant for daily practice. Greenhalgh, Howick, and Maskrey report: “One 2005 audit of a 24 hour medical take in an acute hospital…included 18 patients with 44 diagnoses and identified 3679 pages of national guidelines (an estimated 122 hours of reading) relevant to their immediate care.”\textsuperscript{133} Undoubtedly this statistic would still be much more severe if measured for 2015. Further, given that this statistic is referencing national
guidelines, that is, those guidelines determined after a review of the research literature, this indicates that the original research literature would be exponentially larger and take much more time to digest, making the implementation of textbook EBM wholly impossible just on the basis of its time consuming nature alone.

It is a truism that clinicians face significant time constraints in their daily practice partly due to the nature of medicine and partly due to the increasing demand that they see more patients in a shorter period of time. Either way, these very real time constraints do not allow clinicians time to perform in-depth literature searches or critical appraisals (it is near impossible for a clinician to have time to conduct any level of critical appraisal at the point of care), especially not for every patient they treat. Consequently, clinicians are forced to search for information in the most efficient and easily accessible way, which often means doing basic searches for the most up-to-date critical appraisals, meta-analyses, and/or review summaries as well as consulting other clinicians, pocket guides, etc. This fact is reinforced by Guyatt and colleagues’ observation that “time limitations dictate that evidence based practitioners also rely heavily on conclusions from pre-appraised sources.”

Pointing to this fact is not meant to be a criticism of clinicians. Even a retired clinician or graduate student with nothing but time on their hands would be hard-pressed to read 19 articles a day, everyday. The criticism simply points to the difficulty of practicing textbook EBM under any conditions and in any clinical context. In fact, given the critical analyses of EBM completed thus far and the conclusions that have been drawn, practicing medicine in this way (i.e., accessing pre-appraised sources and other traditional sources) is what a well thought out evidence-based practice should require because it is a practical method for conducting a sophisticated critical assessment of the literature. Accordingly, it is clear that the conception of EBM set forth in 1992 failed to take into account the time

134 Guyatt et al., “Practitioners of Evidence Based Care,” 954-955.
constraints of clinicians and, upon this recognition, needed to revise the set of acceptable evidentiary sources so that clinicians would not need to perform an independent critical appraisal of the literature. It is likely that this oversight of the practical time constraints on clinicians was caused by the overly skeptical stance of EBM proponents regarding appeals to authority. Recognizing the potential harm of appealing to authorities to guide medical decisions encourages a blindly zealous individualism in the original creation and statement of evidence-based clinical practice.

II.I.II Necessary Skills

In Mayer’s brief description of “doers,” he gives the impression that most clinicians fulfill the demands of textbook EBM by searching out and critically appraising the relevant scientific literature. The skill of “critical appraisal,” which Mayer holds to be one of three necessary skills for being able to practice EBM,\textsuperscript{135} and which is also a primary focus of the conception of EBM set forth in 1992, is the focus of “the majority of the chapters in [his] book” and is a “set of skills [that] will help…develop critical thinking about the content of the medical literature.”\textsuperscript{136} Given that it takes the majority of 442 pages to explain and teach the skills necessary for critical appraisal, it is questionable whether these are skills most clinicians currently possess, especially given the fact that a significant number of practicing clinicians completed medical school long before the advent of EBM in the early nineties when statistical methods and clinical trial design were not sufficiently taught as part of medical school curriculum.

Looking to some preliminary research regarding the statistical abilities of practitioners,

\textsuperscript{135} Mayer also lists the skills of Information Mastery, “the skill of searching the medical literature in the most efficient manner to find the best available evidence,” and the process of Knowledge Translation, which involves taking “the results of the information found and critically appraised” and “[applying it] to patient care” (Mayer, Essential Evidence-Based Medicine, 2\textsuperscript{nd}, 10.).

\textsuperscript{136} Mayer, Essential Evidence-Based Medicine, 2\textsuperscript{nd}, 10.
Anderson, Williams, and Schulkin found in their study of statistical literacy in obstetrics and gynecology residents that “respondents performed poorly on two statistical literacy questions, with only 26% (1222 of 4713) correctly answering a positive predictive value question and 42% (1989 of 4173) correctly defining a P value.”¹³⁷ Perhaps even more troubling, they found that 37% of the residents surveyed had “received no formal statistical literacy training in residency.”¹³⁸ Similarly, Windish, Huot, and Green found in their study of internal medicine residents’ understanding of biostatistics and their ability to interpret research results that “seventy-five percent [of residents surveyed] indicated they did not understand all of the statistics they encountered in journal articles, but 95% felt it was important to understand these concepts to be an intelligent reader of the literature.”¹³⁹ These studies, published in 2013 and 2007 respectively, indicate that even newly minted MDs are not acquiring the requisite skills to master the practice of textbook EBM, although it is doubtful that all clinicians could develop such skills, not because of a lack of ability or intelligence but due to the unreasonable standards of expertise textbook EBM requires in these areas.

Lacking statistical skills and knowledge of trial methodology makes analyses and evaluations of clinical trials exceptionally difficult, preventing clinicians from being able to translate what the data are telling them. Additionally, lacking this basic skill set, clinicians are not able to reconcile contradictory studies or evaluate bias or other flaws in trial design. On this point, Singh and Oswald argue:

> Even if clinicians had the time to read all the necessary research literature, many of them would have difficulty sorting through the research design, methodology, and results sections of research articles to determine which treatment would be the

¹³⁸ Ibid., 272-275.
most effective for their patients. For example, clinicians are confounded by research studies on the same topic that often report contradictory findings, some of it probably due to false-positive and false-negative results in small studies.  

The serious lack of skills described by Singh and Oswald reinforces the idea that even if one can alleviate the time constraints clinicians face, at the end of the day a majority of clinicians do not have the skill set to successfully practice EBM, in particular, critical appraisal poses a significant problem.

Guyatt and his colleagues further suggest that part of the problem is the lack of interest of clinicians in learning the necessary EBM skills and their general belief that learning these skills will not improve the practice of medicine. Guyatt and colleagues state:

After a decade of unsystematic observation of an internal medicine residency programme committed to systematic training of evidence based practitioners, we have concluded—consistent with predictions—that not 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 use evidence based summaries generated by others (72%) and evidence based practice 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.”

From these conclusions it appears there is very little that could be done to help the practice of textbook EBM if teaching additional skills to clinicians is not an option. However, an efficient model of evidence-based medicine revolving around the quality and accessibility of pre-appraised evidence where clinicians are allowed to continue to practice with their current skill set would alleviate these problems. Accordingly, it is very likely that clinicians’ lack of and disinterest in the requisite skills for practicing textbook EBM contributed to the acceptance and encouragement of relying on pre-appraised sources in the literature in 1996. However, given the low likelihood that clinicians will be the ones engaging in critical appraisal, it appears a new

---

140 Singh and Oswald, “Evidence-Based Practice”, 130.
141 Guyatt et al., “Practitioners of Evidence Based Care,” 954.
category must be introduced into the taxonomy to account for critically appraising the research literature.

II. II The Need for “Synthesizers”

While the above analyses focused on those doers who engage in the critical appraisal process themselves, it should be mentioned that Mayer also includes “some of [the] “doers” [as] the people who create critically appraised sources of evidence and systematic reviews or meta-analyses.” That is, a subset of doers are the ones responsible for producing systematic reviews or meta-analyses, to be consumed by other practitioners who are less well-versed in statistics and research methodology (i.e., they lack the skills necessary for conducting a “critical appraisal,” or they simply do not have the time, which is most clinicians, to conduct them). However, the contingent of practicing clinicians that will have the time and requisite skills to engage in such activities is small and, more importantly, it is unlikely that they will have comparative expertise to someone whose primary area of expertise is the work of critical appraisal (i.e., expert critical appraisers). Groups like the Cochrane Collaboration exist for the purpose of generating and evaluating meta-analyses and systematic reviews and therefore must represent the portion of doers Mayer is referencing here.

However, the doers whose primarily role is to create pre-appraised sources differ from other doers in an important way. The doers at the Cochrane Collaboration and other institutions like it are only involved in the generation of pre-appraised evidence sources (i.e., critical appraisal) and therefore are not involved in the entirety of EBM practice. The Cochrane Collaboration also employs a number of non-clinicians to generate pre-appraised sources, such

---

142 Mayer, *Essential Evidence-Based Medicine*, 2nd, 11.
as epidemiologists and statisticians. This calls into question whether such individuals can be considered doers or even practitioners of EBM. EBM never explicitly includes expert clinical appraisers in any of its iterations, however, if one considers the pragmatic definition set for by Haynes in 2002, stating that EBM is “a set of tools and resources for finding and applying current best evidence from research for the care of individual patients,” one can arguably include expert critical appraisers as an EBM resource integral to clinical practice, but most likely not as actual practitioners. At this point it is necessary to outline a separate category for expert clinical appraisers, the “synthesizers.”

Because the primary task of these individuals is to find, assess, and present in clinically applicable form the best research evidence, the name “synthesizers” seems appropriate. This category is to include the rare subset of doers (if any) and the set of individuals (including clinicians and non-clinicians) with the requisite training and skills to produce pre-appraised evidentiary sources for use by EBM practitioners. Unique to synthesizers in comparison to those in other categories is that they are not necessarily clinicians and they are not necessarily involved in the later steps of practicing EBM, those beyond critical appraisal (i.e., utilizing what one deems best evidence in patient care). Constructed in this way, synthesizers are solely responsible for generating pre-appraised evidence sources for clinicians. This involves undertaking the arduous task of critically appraising all of the literature relevant to answering a particular clinical question.

To further comprehend the role of synthesizers, an understanding of how pre-appraised sources are created is necessary. The critical appraisal process required for generating a systematic review involves gathering and analyzing all clinical studies related to a particular

---

clinical question. In analyzing each study one must consider the following questions among others: “Was the assignment of patients to treatments randomized?”; “Were the groups similar at the start of a trial?”; “Aside from the allocated treatment were the groups treated equally?”; “Were all patients who entered the trial accounted for? – and were they analyzed in the groups to which they were randomized?”; “Were measures objective or were the patients and clinicians kept “blind” to which treatment was being received?”; and “How large was the treatment effect?”\textsuperscript{144} To answer these questions one must be capable of understanding well the methods and results sections of clinical trials. Not only does this require a nuanced understanding of statistics and trial methodology, it is also a very time consuming process that requires the attention one would give to a full time job. It is not a task to be completed on the spot under serious time constraints. Further, once one has finished assessing the individual trials one must then move on to assessing the group of studies as a whole to determine what conclusions can be drawn from the collection of results one is presented with. If one is only reviewing RCTs then a meta-analysis may be in order and one might also need to generate a forest plot of the individual study results to accurately present their findings. The complexity of these tasks prevents them from being possible to conduct in the clinical setting, however by permitting their use as evidentiary sources, the breadth and quality of evidence used to support clinical decisions is much greater than it would be if depending on the critical appraisal of an individual clinician in the clinic.

Having sufficient mastery over the process of critical appraisal requires significant training and is not something that can be easily learnt in a class or two or through reading a few journal articles. Like most specialized knowledge, it requires mastery and a high level of

\textsuperscript{144} These questions were taken from Oxford University’s Centre for Evidence-Based Medicine’s guide sheet entitled “Critical Appraisal for Therapy Articles.” This guide sheet can be accessed at: \url{http://www.cebm.net/critical-appraisal/} (last accessed September 10, 2014).
expertise. Clinicians are not in a position to acquire such expertise in much the same way a practicing surgeon could not become an allergist by taking a class or reading a couple of articles. In its 1992 construction, EBM unfortunately fails to acknowledge the difficulty and magnitude of the task of critical appraisal and consequently also fails to realize the impossibility of clinicians being equipped to complete such a task. The problem here is not just the limited skills and time clinicians have to practice and master these skills, rather it is a failure to recognize that critical appraisal is a task that requires mastery and specialization in much the same way any specialty in medicine (e.g., rheumatology, orthopedics, surgery, anesthesiology) would. Practicing clinicians cannot be expected to adopt a specialist’s role for conducting critical appraisals, rather critical appraisers must be recognized as constituting a separate specialty requiring professional expertise.

Consequently, it seems then that there is an important contingent of people integrally involved in the practice of EBM – performing its central function: critical appraisal – who do not fall neatly into Mayer’s taxonomy. Given that Mayer is only concerned with clinicians and other health care professionals practicing in the clinic it stands to reason that he has not explicitly added an additional category for those individuals involved in generating pre-appraised sources but not practicing in the clinic. EBM makes similar omissions. However, when one more closely considers the ability of clinicians to act as expert clinical appraisers, one finds that this type of expertise is too vast and complex for a clinician to attain and implement well, not only due to the practical constraints of time and skill mastery, but because critical appraisal should be considered its own specialty (like pediatrics or oncology) because the level of expertise it requires is so high. Consequently, any taxonomy that fails to acknowledge a role for expert critical appraisers is not accurately describing the practice of EBM. In fact, a stronger argument can be made. Given the
centrality of critical appraisal in the 1992 construction of EBM, it cannot be considered an accurate approach to clinical practice. The assumptions of such an approach – that clinical research evidence should be the exclusive source of evidence, that traditional appeals to authority cannot be trusted, and that critical appraisal of the literature is sufficient for resolving clinical questions – narrow it so greatly that at most it can be interpreted as an empirical approach to clinical decision-making (i.e., not an approach to evidence-based clinical practice) specific to making efficacy judgments. Whether EBM proponents explicitly realized that this was the case or not, the insufficiency of such an approach is the most probable justification for why pre-appraised sources became a primary evidentiary source in the 1996 articulation of EBM and also why it allowed for the inclusion of clinical expertise.

At this point, it may be unclear what the profile is of those clinicians encompassed by the doer category and what their function in the practice of EBM is. Given that the subset of doers responsible for generating pre-appraised sources of evidence are subsumed by the synthesizers category, those clinicians properly labeled as “doers” are those who engage in critical appraisal of the literature within their own practice. That is, they consult the scientific literature and critically appraise the studies they read. They are not however responsible for generating and critically assessing pre-appraised sources for others and are able to engage with the literature in this way because they have the requisite skills and the luxury of time. However, after considering the very real clinical constraints on a clinician’s time and acknowledging the turn from textbook EBM in 1996, it can be argued that the doer category is likely very small, although Mayer’s construction of it does serve a heuristic purpose.
II.III. “Users” & EBM (1996)

The second type of practitioners, the “users” according to Mayer, are the benefactors of the synthesizers. Insofar as the synthesizers are responsible for generating pre-appraised sources of evidence, the users are consumers of that knowledge. Users, as described by Mayer, have the skills to carry out effective literature searches and are efficient at determining what evidence will be relevant for patient care. However, the user does not interact with the literature base in the same way that doers or synthesizers would. The skills of critical appraisal appear unnecessary, or at least there appears to be no requirement for clinicians to have a sophisticated command of them, to be a user. That is, users only interact with pre-appraised sources of evidence and therefore do not have, nor do they require, access to individual trial designs and therefore need not evaluate individual studies for bias, etc. However, if one is unable to conduct a critical appraisal of the evidence one is considering or defers that task to others (i.e., synthesizers), one cannot complete the six-step process at the core of textbook EBM on one’s own. In particular, one cannot complete steps 3 or 4. Users are capable of completing the other steps in the EBM process, that is, crafting a question, locating the relevant sources, and determining how that evidence would apply to their patients, however, they do not engage in critically appraising individual studies and therefore have to rely on the products of synthesizers to complete these steps.

---

145 In Mayer’s text he considers them to be benefactors of the doers, however, given the introduction of synthesizers as the group containing those who generate all pre-appraised sources, it becomes necessary in this chapter to talk about users as beneficiaries of the synthesizers for consistency’s sake.

146 The steps involved in this process include: 1. Craft a clinical question that includes the patient, the intervention, the comparison, the outcome of interest, and the time frame (often called the PICO or PICOT formula); 2. Search the medical literature for those studies that are most likely to give the best evidence. (This step requires using good searching skills using medical informatics.); 3. Find the study that is most able to answer this question. Determine the magnitude and precision of the final results; 4. Perform a clinical appraisal of the study to determine the validity of the results. Look for sources of bias that may represent a fatal flaw in the study; 5. Determine how the results will help you in caring for your patient; 6. Evaluate the results of applying the evidence to your patient or patient population. This iteration of the six-step process was taken from: Mayer, Essential Evidence Based Medicine, 2nd, 14-15. Articulations of the six-step process vary from author to author but all assert the same six general steps.

147 See footnote 33 for the complete six-step process.
Most health-care workers will spend a greater part of their time functioning as "users" of the medical evidence. They will have the skills to search for the best available evidence in the most efficient way. They will be good at looking for pre-appraised sources of evidence that will help them care for their patients in the most effective way.\(^{148}\)

As one can see from Mayer’s description, EBM users practice a form of EBM motivated by efficiency and effectiveness, most closely resembling the 1996 articulation of EBM. Guyatt and colleagues, in an essay addressing whether all clinicians need to be well versed in critical appraisal skills, explain that

Trainees who are less interested in evidence-based methods develop a respect for, and ability to track down and use, secondary sources of pre-appraised evidence (evidence based resources) that provide immediately applicable conclusions. Having mastered this restricted set of skills, these trainees (whom we call evidence users) can become highly competent, up to date practitioners who deliver evidence based care.\(^{149}\)

Guyatt and his colleagues, many of whom were integral to the founding of EBM, seem to be expanding the practice of EBM to include those clinicians who never interact directly with the evidence base (i.e., the original clinical studies), a description that fits neatly with Mayer’s concept of users. Their engagement with the literature is primarily strategic, not critical. This is not meant to be a criticism of the users. The textbook practice of EBM by doers is time consuming, as was established above, and when one is caring for 30-50 patients a day in the clinic it is unreasonable, in fact impossible, to expect any clinician to complete a thorough critical appraisal of the literature for each and every patient they treat. The constraints and demands of the clinic make it necessary to practice a more abbreviated form of EBM, appealing to the authority of pre-appraised sources when making treatment decisions.

\(^{148}\) Mayer, Essential Evidence-Based Medicine, 2\(^{nd}\), 11 (italics mine).

\(^{149}\) Guyatt et al., “Practitioners of Evidence Based Care,” 955.
However, even though users defer to the pre-appraised sources of synthesizers, the generation of such sources by more than one authority introduces a new problem for the profile of users. That is, there is a possibility that the results of these sources will conflict and therefore one would need to conduct a critical appraisal of the pre-appraised sources to determine which source was most accurate and if any of the sources utilized flawed methodology or allowed for inappropriate biases. As users, they would not have the skills to solve such a problem through critical appraisal. Further, even if there is only one pre-appraised source to answer a clinical question, one can still argue that it is necessary to conduct a critical appraisal to ensure the reliability and accuracy of the methodology used to generate it. Again, this moves the task of critical appraisal back to clinicians. Such a critical appraisal would involve answering the following questions: “What question (PICO) did the systematic review address?”; “Is it unlikely that important, relevant studies were missed?”; “Were the criteria used to select articles for inclusion appropriate?”; “Were the results similar from study to study?”; and “How are the results presented?”150 Again, the expertise required to answer these kinds of questions is not within the purview of most clinicians, especially not users. Consequently, it appears that the provision of pre-appraised sources for clinicians is only helpful in narrowing down the number of sources a clinician needs to appraise critically, but it does not relieve the clinician of the duty of critically appraising sources, a task better suited to an expert in critical appraisal.

Given the complexity of conducting critical appraisals and the necessity of clinicians having a reliable authority to appeal to to be able to carry out their daily practice, it may be necessary for clinicians to consider the pre-appraised sources of institutions like the Cochrane Collaboration as authoritative and not in need of further critical appraisal, effectively removing

---

150 These questions were taken from Oxford University’s Centre for Evidence-Based Medicine’s guide sheet entitled “Systematic Review Appraisal Sheet.” This guide sheet can be accessed at: [http://www.cebm.net/critical-appraisal/](http://www.cebm.net/critical-appraisal/) (Last accessed September 10, 2014).
this task from clinicians’ responsibilities within EBM. If a clinician comes across conflicting pre-
appraised sources answering the same clinical question this may be a warrant to consult an
expert critical appraiser, but it cannot require further action on the part of the clinician, whose
skepticism is justified, beyond having one use their clinical judgment to resolve the issue for the
current patient they are treating. Epistemically, the clinician would have competing evidence for
the truth of the matter and therefore would be justified in trusting either source given they are
both produced by known reputable authorities.

A further concern for pre-appraised sources is the sheer volume that exists, as well as an
inability for them to be produced fast enough to encompass all of the current research literature.
This problem is unsurprising as it is the consequence of the original research literature being so
vast. In a 2004 article, Guyatt and colleagues indicate that even if clinicians turn to relying on
pre-appraised sources of evidence in their practice of EBM, it still remains an impractical and
inefficient process for vetting the evidence provided by the literature base. This is because there
are so many clinical trials that even with institutions like the Cochrane Collaboration working
full-time to provide easy access to pre-appraised evidence sources there is still a surplus of
evidence that isn’t captured in the meta-analyses and review summaries produced. Guyatt and
Haynes explain:

Evidence based medicine, still young, faces challenges in integration into clinical
practice. The process of producing relevant evidence through high quality
research will continue indefinitely, requiring considerable investment by funding
agencies all over the world. The process of summarizing that evidence is
daunting. Estimates based on current rates of publication of randomized trials and
completion of systematic reviews indicate that it would take reviewers until 2015
[11 years] to produce the 10,000 Cochrane reviews required to summarize
existing evidence.\textsuperscript{151}

\textsuperscript{151} G.H. Guyatt, D. Cook, and B. Haynes, “Evidence Based Medicine Has Come a Long Way. The Second Decade
Will Be As Exciting As the First,” \textit{British Medical Journal} 329 (2004): 990-991.
The situation Guyatt and Haynes describe is clear. We have no problem producing randomized control trials to provide the evidence base we are looking for, but the challenge comes when trying to relay that information to clinicians. The article, published in 2004, projects that it would take until 2015 to complete systematic reviews reflective of the evidence base accumulated up until 2004, a total of eleven years time. Additionally, this problem is exacerbated when one takes into account that the numbers provided by Guyatt and colleagues encompass only RCTs. If one includes all other clinical trial methodologies (e.g., observational studies) the time required to digest and package all of the evidence for clinicians would multiply exponentially, especially when one considers the difficulty of weighting evidence produced by different methodologies against one another to make final recommendations. Further, the fact that 10,000 Cochrane reviews are necessary to summarize the literature from 2004 indicates that referencing pre-appraised sources still does not manageably narrow the pool of evidence clinicians must consult.

The consequence of these circumstances is that many clinicians, overwhelmed by the surfeit of evidence, justifiably choose not to practice EBM in accordance with its 1996 articulation some or all of the time and those who do attempt to practice it may be prevented from doing so because the evidence base is so unwieldy it is impossible to gather and determine what best evidence is, even if one isolates its definition just to RCTs and pre-appraised sources.

This circumstance highlights the importance and necessity of providing sources of evidence to clinicians that can be easily consumed and do not require further appraisal. That is, clinicians are in a circumstance in which they need to be able to appeal to an authority (e.g., the Cochrane Collaboration) without further need of critical appraisal, and to be able to act on that evidence source as if it were reflective of scientific truth. It is probable that for this reason, as well as the fact that pre-appraised sources are still significantly numerous, EBM proponents
moved away from requiring direct contact with pre-appraised sources for one to practice evidence-based medicine. The beginnings of this change occurred as early as 1995 when Sackett and Rosenberg recognized “accepting evidence-based practice protocols,” “receiving advice from a respected teacher (who has learned EBM)” and “being visited by a non-commercial ‘detailer’ (who is informing and encouraging us about specific evidence-based ways of caring for patients rather than exhorting us to prescribe specific drugs)” as ways to keep “up-to-date.”\footnote{Sackett and Rosenberg, “Need for Evidence-Based Medicine,” 623.} So long as the authoritative sources one consults are informed by the best clinical research, one is considered to be practicing EBM on the 2002 model.\footnote{The 2002 model is referenced here rather than the 1996 model or before because in 1996 the theory was still very much focused on the PICO/PICOT progress and its approach to clinical practice was not nearly as pragmatic as it is in the 2002 model. Therefore, even though there are strains of this thought as early as 1995 it is most appropriate to associate it with the more pragmatic model of EBM in 2002.} Once again it appears that critical appraisal and a narrow definition as to what can be considered a high quality evidentiary source and the practical difficulties they cause is responsible for motivating a shift in EBM theory. This realization is most likely what prompted EBM proponents to reevaluate traditional evidentiary sources and how traditional clinical practice is structured, ultimately motivating a shift to an EBM model focused on infusing the research literature into all evidentiary sources, not just typical pre-appraised sources (e.g., systematic reviews).

II.IV “Replicators”

Mayer calls the third type of EBM practitioners “replicators.” He identifies replicators as those “who simply accept the word of experts about the best available evidence for care of their patients.”\footnote{Mayer, Essential Evidence-Based Medicine, 2nd, 11.} It is unclear from Mayer’s brief description how exactly he conceives of a replicator. Given the minimal constraints he places on the evidentiary sources they consult – really the only
qualification being that it be based on “best available evidence,” notice that research evidence is 
not here required – it appears that Mayer’s replicators are best understood as reflecting the 2002 
EBM model. Under this model the evidentiary constraints are dramatically loosened in 
comparison to previous articulations of EBM practice, making it possible for both evidence-
based and traditional evidentiary sources to be considered in the clinical decision making 
process. Clinical expertise and clinical experience also enjoy more prominent roles, allowing for 
the acceptance of a wider range of appeals to authority. Given the focus of the 2002 model on 
infusing medicine with the best research evidence, it is likely that most traditional sources, like 
textbooks, are now evidence-based, at least to some degree. However, it also remains possible 
that some evidentiary sources may not be based on research evidence at all. After considering 
some research evidence regarding the habits of clinicians when accessing evidentiary sources, it 
will be made clear that the most important function of EBM at this point in time, if it wants to 
increase the use of best research evidence in practice, is to develop easily accessible traditional 
resources based on the best research evidence.

There are three possible interpretations of Mayer’s replicators – possibly more – which is 
fitting given the practice variations, permitted under the 2002 model due to its recognition of 
traditional practice and its evidentiary sources being essential to clinical practice. The first 
interpretation regards replicators as those clinicians who strictly abide by clinical practice 
guidelines or evidence-based health policies at the hospital, state or local level. The “experts” on 
this interpretation would usually be those institutions responsible for crafting the relevant 
guidelines. As such, a replicator would be a clinician who does not consult the original research 
literature for guidance, but instead defers to those guidelines most relevant to the clinical 
decision at hand. A clinician is well justified in deferring to such guidelines as they are
authoritative and have binding force on the clinician. The further hope is that they are sufficiently evidence-based, although not all current guidelines are.

The second possible interpretation is a clinician who is mindful of guidelines but also seeks out additional sources, such as pocket references, other clinicians, etc., to satisfy oneself that the guideline is in fact supported by good scientific rationale. On this hybrid approach the clinician still appreciates the authority of the guidelines but also seeks other more traditionally authoritative (i.e., regarding their expertise and/or experience) sources to further satisfy oneself that the recommendation is appropriate.

The third possible interpretation refers to those clinicians who do not intentionally seek out evidence-based sources (i.e., RCTs and pre-appraised sources) and instead consult only traditional sources, such as quick references, textbooks, and other clinicians. These sources are presumably not meant to constitute the entirety of the evidence base if other, better quality evidence exists, however, as it will be shown, that is not always taken into account. On this interpretation, the “experts” refer to those authorities EBM initially took issue with – that is, deferring to the expert opinion of older and/or more experienced physicians and referencing sources presenting pathophysiological understandings of treatment and disease (e.g., some textbooks) – what Mayer describes as the “traditional hierarchical system of medical practice and the acceptance of the scientific method as the governing force in advancing the field of medicine.”¹⁵⁵ The quality of this type of replicator’s evidence will depend on the degree to which the traditional sources one consults have been informed by the research literature.

There are convincing data that demonstrate the frequency with which clinicians choose not to seek out “evidence-based sources” (i.e., those methodologies appearing at the top of the

¹⁵⁵ Mayer, Essential Evidence-Based Medicine, 2nd, 10.
hierarchy) and instead defer to senior physicians, textbooks, etc. or they seek out evidence-based sources and when they find that these conflict with their clinical judgment, they reject them. Understandably, a number of clinician’s evidence-seeking habits develop based on what evidence is quick and easy to attain and as a result of the research literature being too expansive to manage. As McCord and colleagues observed in their study of residents in family medicine,

> Despite being trained to use traditional EBM methods, our residents most often used authoritative sources to answer their clinical questions, principally, direct physician-to-physician consultation. These authoritative sources may or may not be evidence-based depending on the EBM skills of the consulting physicians as well as the accuracy and reliability of the electronic programs and textbooks used.\(^\text{156}\)

The ease of accessibility that comes with consulting another clinician or accessing a program on one’s PDA makes them, understandably, the most preferred sources for clinicians. When looking at the breakdown of sources consulted in the study, McCord and colleagues found that “residents were observed…to use attending physicians, PDA programs (mostly ePocrates), and textbooks as 87% of the sources they consulted.”\(^\text{157}\) This means a mere 13% of the time clinicians consulted the research literature looking either at RCTs or pre-appraised sources. This statistic becomes even more telling when one realizes that the residents observed all underwent training to make them successful practitioners of EBM. However, as Ramos, Linscheid, and Schafer report in their 2003 study\(^\text{158}\) conducted to determine whether residents trained in EBM consulted “evidence-based sources” more frequently than practicing physicians, “resident and faculty

\(^\text{156}\) McCord et al., “Answering Questions,” 300-301.
\(^\text{157}\) Ibid., 300.
\(^\text{158}\) According to Ramos, Linscheid, and Schafer, their study “explored the assumption that current residents and their faculty frequently apply the principles of EBM.” “The study involved physicians associated with the University of California San Francisco-Fresno (ECSF-Fresno) Family Practice Residency Program, a community-based residency program affiliated with the University of California, San Francisco.” “We invited 40 physicians to participate. All agreed to do so, but we could not accommodate one resident and three faculty physicians due to scheduling conflicts. We spent 125 hours observing 11 faculty member and 25 residents, 8 in their third year, 11 in their second year, and 6 in their first year…The resident group together was 84% male and the faculty 64% male. The mean age of residents was 36, and the mean age of faculty was 41” (K. Ramos, R. Linscheid and S. Schafer, “Real-Time Information Seeking Behavior of Residency Physicians,” *Family Medicine* 35 (2003): 257-8).
family physicians have many clinical questions but rarely use evidence-based information sources to answer these questions.”

From this conclusion it appears that EBM training is not a reliable predictor of whether a clinician will practice EBM. In fact, it appears that even those being inducted into medicine at a point when practice is meant to be wholly evidence-based still are not practicing EBM. Singh and Oswald further reinforce the veracity of this claim by summarizing the state of EBM practice. They hold:

Given the impossibility of keeping abreast of the research literature, this [has] led (at least in part) to wide variations in clinical practice…surveys show that up to 40% of clinical decisions made in academic medical centers are without any research support, suggesting that clinicians rely more on their own experience with a limited number of patients, academic detailing from pharmaceutical representatives, or workshops by experts than on systematic reviews of the current literature (Greenhalgh, 2001).

The evidence presented here by Singh and Oswald points to a practice of medicine that is not wholly based on the research literature and therefore a practice of medicine, EBM still considers plagued by expert judgment and other forms of “dogmatic medicine.” Additionally, with “wide variations in clinical practice” it becomes impossible for EBM’s purported benefit of achieving standardized practice (i.e., that patients will receive similar care from different health practitioners) to be realized.

Recognizing the severity of the situation and a need for change, Ramos, Linschield, and Schafer suggest, “EBM curricula should acknowledge the time limitations of the clinical setting and help physicians become familiar with convenient and available evidence-based sources that yield speedy answers, and explore systematic methods of resolving unanswered questions.”

---

160 Singh and Oswald, “Evidence-Based Practice,” 130.
161 It should also be noted that even if there were not wide variations in clinical practice it would remain the case that practice could not be standardized because different clinicians sometimes interpret clinical evidence (even best evidence) differently and the specific physiology, co-morbidities, and preferences of the patient will require differences in treatment for the same diagnosis.
These statistics undoubtedly make clear the need for EBM to move beyond requiring critical appraisal (either of the original sources or pre-appraised sources) in the clinical context to an evidence-based clinical practice where the best research evidence is present in those traditional sources they find easily accessible and trustworthy.

The 2002 model of EBM, a dramatic shift from the approaches to EBM that came before it, goes a great distance in achieving the goal of producing easily accessible sources of evidence. However, while EBM’s model of clinical practice no longer requires critical appraisal by clinicians in the clinical context, it still remains the case that the PICO/PICOT process is taught and critical appraisals are still produced. Additionally, clinicians continue to practice in the most efficient way possible, not necessarily concerning oneself with what the “best research evidence” actually is and simply trusting the guidance of those sources one considers authoritative. While it is good that the clinical model has shifted, the continuation of these other practices makes it exceedingly difficult for EBM to achieve its goal of standardized practice; as it encourages various forms of clinical practice. This indicates that even though the EBM literature evolved to the point of considering EBM a compliment to medicine, its previous instantiations have become permanent artifacts in the clinic. With all of these instantiations of EBM bleeding into the current model, the problem arises as to what qualifies as best evidence and whether consulting anything less than best evidence is a violation of EBM. Moreover, how is one to know what is the best evidence? If best evidence can only be constituted by those research methodologies appearing at the top of the evidence hierarchy (i.e., RCTs and pre-appraised sources) and one needs to know which are the best, then the devastating consequence of this will be that most sources clinicians use in regular clinical practice can never be considered sufficient for making treatment decisions. Given the widely accepted rhetoric that has sprung up regarding EBM’s categorical prioritization
of RCTs, it stands to reason that the only evidence one should consider on such a view is that generated by RCTs and pre-appraised sources, and as such the aforementioned consequence holds. Alternatively, even if it is the case that all forms of evidence, including guidelines, textbooks, etc. (evidence-based or not) can be considered part of the evidence base eligible for being included as best evidence or as good enough evidence, it is not obvious how this reconciles with Mayer’s and most EBM proponents’ focus on critical appraisal as essential to EBM – a necessary tool only if one is consulting research evidence.

After considering Mayer’s taxonomy, the likely interpretations of his categories, and the addition of a new category (i.e., synthesizers), it is clear that the practice of EBM is very messy and does not admit of a universal definition or practice in practice. This is unsurprising when one considers the many ways in which the standards of EBM lead to impossible consequences for practice. The conclusions to be drawn from these sections are that most well intentioned clinicians will attempt to be users, relying on the products of synthesizers. Another contingent of clinicians who are disinterested in, overwhelmed by the practice of, or who recognize the foundational problems with EBM will function as replicators. Only a very small percentage of clinicians will be able to function as doers or synthesizers due to a lack of training and skills and/or constraints on their time. Additionally, there is an argument to be made that the work of critical appraisal is better left to professional critical appraisers (e.g., epidemiologists and statisticians) and therefore clinicians should not attempt to inhabit the role of synthesizer.

III. Taking Stock

In this section an evaluation of whether the current practice of EBM goes any distance in accomplishing the six widely recognized goals claimed for evidence-based practice will be
undertaken. The six goals to be assessed are: 1) minimizing the inverse relationship between healthcare costs and patient outcomes; 2) eliminating clinicians’ deference to an authority; 3) discouraging the use of clinical experience and clinical expertise; 4) standardizing medical practice; 5) improving patient outcomes; and 6) better integrating clinically relevant research into medical practice. Each of these purported goals will be assessed in this section however the conclusions reached will demonstrate that goal #6 is achieved, goals #1-#4 are not, and the status of goal #5 is indeterminate due to a current lack of information. It should be noted that goals #1 and #5 were not previously included in the analyses of chapter two because they are practically oriented goals that cannot be assessed on the basis of a theory or model of clinical practice.

To begin, there is no evidence to suggest that the move towards evidence-based practice has had any substantial financial benefits for patients, that is, clinical care has not decreased in cost while at the same time increasing patient outcomes. There are currently no studies directly taking up this question, however the continued rise in healthcare costs in all areas suggests that EBM has not been successful in decreasing the cost of care. To substantiate this claim, consider a study conducted by Moses and colleagues, indicating that healthcare expenditures have steadily risen in the United States since the 1980s. All major areas of expenditure have increased, including administrative costs, professional services, healthcare facilities, and especially the costs of prescription drugs and medical devices, according to their findings.\textsuperscript{163} While analyses of this kind are complicated and need to be sensitive to many factors, it remains safe to say that EBM has not had an overwhelming effect on the cost of care, although further research would be required to make any detailed claims of this nature.

Second, it is abundantly clear that EBM has not succeeded in eliminating the need of clinicians to defer to an authority in clinical practice. It also has been noted that this goal is

wholly inappropriate for any realizable account of medical practice. It may be the case that the quality of the authorities referenced have changed, however, clinicians still must defer to authorities other than themselves to fulfill their evidential needs in the clinical context because there are no practical means available to them to escape this circumstance. This claim is supported by the basic fact that the volume of research evidence provided by the literature relevant for clinical practice is too great for any individual clinician to appraise on her own. Even with the advent of pre-appraised sources like meta-analyses and systematic reviews, the institutions and collaborations responsible for producing them still cannot command the breadth of the literature. Moreover, clinicians cannot have the expertise of professional appraisers, that is, the experts. This makes necessary the installment of authorities whose primary occupation is appraising and packaging the literature in an easily accessible and reliable form so clinicians can remain up to date on the current literature.

Third, clinical experience and clinical expertise, despite EBM’s best initial efforts, have remained primary evidentiary sources for clinicians. As was mentioned above, clinicians overwhelmingly continue to reference their own experience or the opinions of other trusted clinicians to inform their clinical decision-making. However, given the 2002 model of EBM, and its reunion with the components of traditional medicine, this may not be as much of a problem for EBM as it once was. Additionally, given that EBM has managed to infuse the best clinical research into a wide variety of traditional sources and have also generally made clinicians more aware of the clinical evidence, appeals to clinical experience or clinical expertise may not be as untrustworthy as they once were. That is, the general knowledge base of clinicians has been better informed by the current research literature, meaning that a recommendation from another
clinician may in fact be more reliable than it once was. However, on a strict interpretation of this goal, EBM has not achieved it.

Fourth, the claim that evidence-based medicine assists in standardizing medical care for patients is demonstrated to be false when one recognizes the variety of ways in which clinicians currently practice EBM (i.e., the existence of doers, users, and replicators), the variance in the evidence sources they consult, and the non-standardized guidelines employed in creating pre-appraised sources. Depending on the literature a clinician is exposed to it is entirely possible that he would reach a different conclusion regarding the best treatment than his colleague examining different components of the same literature. It is also the case that two clinicians looking at identical literature sources may reach different conclusions on the basis of professional or personal biases. However, these circumstances are not unique to EBM. In any paradigm of medicine there will be variance in treatment decisions because even if all clinicians are referencing the same evidence in their treatment decisions it will always be the case that the subjective judgments of both the clinician and the patient will play an influential role. As such, standardizing medical practice remains an impossible goal for EBM.

Fifth, regarding improvements in patient outcomes, evidence does not yet exist indicating whether EBM has had a significant impact in this area. Such research projects would be difficult to undertake because it would require one to: control for medical advances brought about by technology and pharmacology, the changes in the general health of the population, advancements in preventative medicine, as well as delineate between practitioners who are and are not practicing EBM. Controlling for these variables would be difficult at best and most likely impossible. Therefore, there is not much one can say to definitively evaluate this proclaimed goal of EBM.
Lastly, EBM’s aim of integrating clinically relevant research into daily practice has overwhelmingly been achieved. It may be the case that clinicians do not have direct contact with the research evidence in the manner hoped for, however, it certainly has become common practice to support medical claims and warrants with the most recent clinical research whether through direct contact with this research or through its infusion in other traditional sources. Medical textbooks, journals, guidelines, policies, etc. are all evidence-based in current medical practice. The need to support one’s claims with current research has become undeniable. The level of success in achieving this goal is also demonstrated by the fact that the reliance on research evidence for establishing professionally oriented claims championed by EBM has infiltrated many other disciplines. A quick search of the internet will produce a variety of pages dedicated to evidence-based education, psychology, computing, apparel, business, dentistry, and many more. So, if nothing else, EBM has achieved its aim of bringing current research to the forefront of medical practice as well as many other disciplines.

From the analyses above one can see that current evidence-based practice does not closely resemble the initial theoretical construct issued in bringing evidence-based medicine to the forefront of medical practical in 1992. The appropriation of EBM in clinical practice also has failed to achieve most of the goals set forth by the EBM movement making the effectiveness of the “new medical paradigm” questionable at best. However, medical practice has been greatly affected by the incorporation of current research into practice and in this regard EBM appears to have achieved its aim.
CHAPTER 4
ASSESSING RESEARCH EVIDENCE OF CLINICAL EFFECTIVENESS: THE EVIDENCE HIERARCHY

I. Introduction

The previous two chapters examined evidence-based medicine’s approach to clinical practice. It was established that historically three distinct theories of clinical practice were constructed in the medical literature, always changing to accommodate the shortcomings of the previous theory. The strongest and most consistent thread running through each of these theories, central to each definition of EBM set forth, is the idea that clinical research evidence or “best evidence” needs to be incorporated into medical decision-making and that clinicians should guide their evidentiary choices by employing the rules of evidence.164

Best evidence is never explicitly defined in the literature because it is meant to be a contextual concept, encouraging clinicians to locate the best available evidence for the particular decision at hand. While this sounds like a pragmatic approach, it has amply been made clear by proponents of EBM that RCTs are the gold standard of evidence and, if available, should be preferred over all other methodologies, effectively making it a categorical approach. The rules of evidence, originally laid out by Sackett in 1989 and now institutionalized and oft represented by the evidence hierarchy, are meant to provide a ranking schema for determining the best available evidence by which clinicians can assess different sources of evidence based on the methodologies that produced them (e.g., RCTs or observational studies). The rules of evidence, encoded as the evidence hierarchy, have been the focus of much contention in the philosophical EBM literature. In particular, EBM critics, claiming that the attribution of internal and external validity to RCTs is problematic, have challenged the gold standard status of RCTs.

164 In chapter 2 Sackett’s 1989 rules of evidence were introduced. See: Sackett, “Rules of Evidence,” 2S-4S.
In this chapter, EBM will be analyzed and critiqued as an approach for assessing research evidence of clinical effectiveness – the proclaimed purpose of the evidence hierarchy.\(^{165}\) To begin, the evidence hierarchy will be explained in more detail with a few examples given.\(^{166}\) The two justifications of the hierarchy’s ranking schema, largely tied to the proclaimed methodological superiority of RCTs, will be set forth and analyzed. The first justification, which claims that RCTs have a superior ability to identify causal relationships between treatments and their effects, will be rejected on the grounds that causal relations are only better identified by *ideal* RCTs, which do not exist in the clinical research setting. The second justification, which holds that RCTs produce less biased results (i.e., better control of known and unknown confounding factors), will also be rejected on the grounds that randomization does not achieve the elimination of bias in the ways expected by EBM proponents and, in addition, it is also not the only methodological feature capable of controlling for some types of bias (an alternative would be masking\(^{167}\)). After addressing these issues, it will be argued that even if critics of the hierarchy are wrong in their assessment of RCTs, further concerns persist regarding their external validity (i.e., “whether the results can be reasonably applied to a definable group of

---

\(^{165}\) Evidence hierarchies and rules of evidence exist for a variety of different clinical questions. There are some differences based on the particular clinical questions being considered (e.g., diagnostic, prognostic, etc.). This chapter will strictly focus on the clinical efficacy of treatments and interventions because it is the focus of the bulk of the literature.

\(^{166}\) It should be noted that there is no universally accepted hierarchy of evidence. There are a number of depictions of the evidence hierarchy found in the EBM literature. For the most part the ranking schemas they employ are similar, but they are not identical. Common among all of them is the superior status of RCTs, however some put meta-analyses and systematic reviews above RCTs because they are generated from the pooling of RCTs or a critical appraisal of RCTs for a particular topic. Additionally, some hierarchies include clinical expertise, clinical experience, and pathophysiological rationale at the bottom of the hierarchy whereas others do not. Lastly, observational studies always appear directly under RCTs, but the different types of observational studies are sometimes ranked in different orders. For an extensive database of different EBM evidence hierarchies please visit Christopher Blunt’s page at: [http://cjblunt.com/hierarchies-evidence/](http://cjblunt.com/hierarchies-evidence/) (last accessed May 10, 2015).

\(^{167}\) The term masking is being used here to indicate what has traditionally been known as blinding. There has been much debate in the medical literature over which term is most appropriate. Blinding has a 200-year history of use, however masking avoids patient confusion in ophthalmologic research settings and is also easier for ESL speakers to understand. The author values practical considerations over tradition and therefore will employ the term masking.
patients in a particular clinical setting in routine practice.\textsuperscript{168}, which should be taken into consideration when assessing the epistemic superiority of different trial methodologies. In particular, it will be shown that RCTs are largely incapable of providing evidence applicable for clinical use due to the methodological features of RCTs (e.g., run-in periods and exclusion criteria) that narrow the subject pools so significantly that they cannot provide useful information about the chosen target population(s) (i.e., the patient population that will actually receive the intervention). Accordingly, it will be asserted that concerns over the internal validity of trial methodologies should not be the only epistemic factor considered in the hierarchy’s ranking schema.

II. Evidence Hierarchies

The evidence hierarchy (and the rules of evidence that underpin it) seeks to categorically rank the different research methods employed in clinical research. As Kirstin Borgerson observes, “ranking research methods is not at all intuitive, nor is such ranking widely practiced by scientists.”\textsuperscript{169} While this may be the case, the attempt to do so in medicine, an applied science, is largely motivated by the need for clinicians to be able to rapidly access and appraise evidentiary sources in the clinical setting. Accordingly, it is asserted that knowing which methods (supposedly) provide the highest quality evidence can help clinicians navigate the literature more efficiently. In the words of Ross E.G. Upshur, “levels of evidence…attempt to standardize and provide clinicians with cogent rules to appraise published research, determine its validity and

\textsuperscript{168} P.M. Rothwell, “External Validity of Randomised Controlled Trials: “To Whom Do the Results of this Trial Apply?”” \textit{The Lancet} 365 (2005): 82.

summarize its utility in clinical practice." The importance of clinicians having easy access to high quality evidentiary sources in the clinic cannot be overstated and has motivated many of the advances in EBM as well as the rapid rise of medical information technologies.

In general, the evidence hierarchy follows a ranking schema similar to that described by the American Dental Association. Their ranking is as follows.

- Randomized controlled clinical trials, non-randomized controlled clinical trials, cohort studies, case-control studies, crossover studies, cross-sectional studies, case studies or, in the absence of scientific evidence, the consensus opinion of experts in the appropriate fields of research or clinical practice. The strength of the evidence follows the order of the studies or opinions listed above.

While many often refer to the evidence hierarchy, in practice there are well over 100 different hierarchies in circulation. Some differ in the methodologies they include or how they rank different types of observational studies, but in general all hierarchies adhere to three central principles. Jeremy Howick concisely states them as follows.

1. Randomized trials (RCTs), or systematic reviews of many randomized trials, generally offer stronger evidential support than observational studies.

2. Comparative clinical studies in general (including both RCTs and observational studies) offer stronger evidential support than “mechanistic” reasoning (“pathophysiological rationale”) from basic sciences.

3. Comparative clinical studies in general (including both RCTs and observational studies) offer stronger evidential support than expert clinical judgment.

From these principles one can see their connection to the original statement of EBM in 1992, which mandated a lower epistemic status for clinical expertise and pathophysiological rationale and prioritized the use of clinical research (i.e., RCTs and observational studies). Additionally,
after a quarter of a century, the rules of evidence set forth by Sackett in 1989 still hold firmly by privileging RCTs over observational studies.\textsuperscript{174}

Figure 1

![Evidence Hierarchy Diagram]

Figure 1 is a typical pictorial representation of the evidence hierarchy. One can see that the highest-ranked methodologies appear at the top of the pyramid and the lowest ranked methodologies appear at the bottom. Additionally, one should note that the top half of the pyramid only contains RCTs and those pre-appraised sources (i.e., meta-analyses and systematic reviews) generated from RCTs. Background information and expert opinion appear on the bottom of the hierarchy, however it is not uncommon for them to be separated from the systematized methodologies ranked above them or omitted completely. As one can see, this hierarchy accurately reflects the principles set forth by Howick.

One should be careful to note that the relationships identified between methodologies by Howick’s principles hold only generally. This is the consequence of recent efforts to loosen the

\textsuperscript{174} See: Sackett, “Rules of Evidence,” 2S-4S.
categorical ranking of the evidence hierarchy by permitting more nuanced rankings depending on one’s clinical question. However, even these more nuanced rankings put RCTs at the top of the evidence hierarchy and favor its methodological features (i.e., randomization, masking, and the use of controls). While this is a positive development, it remains the case that RCTs are still by and large held to be the gold standard for a number of reasons, not the least of which is “a widely held belief that the RCT is the most scientific and rigorous study design available (i.e., it has high internal validity). Any study using the RCT design is considered superior to any study not using this design.” This belief has been reinforced by the biases of funding agencies and the policies of scientific journals. Grossman and Mackenzie report that “one has much more hope of getting a bad RCT funded than any caliber of, say, observational study, both because of the opinions of funding bodies and, less directly, because of the publication policies of journals.” Given the widespread acceptance of the superiority of RCTs, it is necessary to ensure that the methodological and epistemic virtues attributed to them actually hold.

III. Justifications, Assumptions and Critique

EBM proponents assert RCTs deserve their gold standard status for two primary reasons (i.e., these reasons justify the ranking schema of the hierarchy): 1) they are more effective at isolating causal relationships and 2) they are better at minimizing bias. Proponents argue that the RCT is the only methodology capable of isolating causal relationships. Accordingly other methodologies simply indicate correlations, or so the claim goes. Regarding the minimization of bias, other

---

175 Howick has been instrumental in such efforts at the Oxford Centre for Evidence-Based Medicine (OCEMB). He was responsible for the 2009 update to the OCEMB. “Levels of Evidence,” accessed June 16, 2015, http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/.
177 Ibid., 517.
methodologies, like observational studies, are considered capable of minimizing bias but not to the extent that an RCT can and therefore they are ranked accordingly. If these claims can be disproven, the epistemic justifications provided for the evidence hierarchy’s ranking schema cease to exist and EBM proponents are left only with bold assertions about their preferences for certain methodologies.

Before examining the epistemic justifications set forth by EBM proponents it is necessary to note that they are related through their reliance on randomization and controlling for confounding factors to establish their claims. In the case of isolating causal relationships, all confounding factors must be controlled for between the treatment and control arms of a study (through randomization), otherwise one cannot determine whether an effect (positive patient outcome) is the result of the cause of interest (the experimental treatment) or another cause/combination of causes (e.g., co-morbidities and/or gender). In the case of minimizing bias, confounding factors, such as selection bias and ascertainment bias, are controlled for through randomization to ensure one’s results are objective, that is, the preferences or values of those involved in the trial do not skew the data.

III.1 Isolating Causal Relationships

Taking causal claims first, EBM proponents assert “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.””\textsuperscript{178} According to EBM proponents, lower ranked studies, like observational studies, only get at “mere correlation” because - even though treatment effects in the target population are observable, known confounders are controlled for,

\textsuperscript{178} Borgerson, “Valuing Evidence,” 221.
and pathophysiologial rational along with large effect sizes are used to minimize the possibility of unknown confounders - the lack of randomization allegedly makes it impossible to control for all unknown confounders, making it impossible to establish a causal claim. On the other hand, RCTs are supposed to be “able to make causal claims based on their methodological structure. Randomization provides closely matched groups and controlled manipulation of variables ensures that one group differs from the other by the variable of interest alone. Thus, any differences recorded between the groups have to be due to the variable of interest. This is what makes the claim causal.” Essentially, the methodology of the RCT ensures that the study is designed in such a way as to isolate the causal claim of interest, specifically achieved through randomization and the use of a treatment and control group (the experimental treatment being the only variable factor between the two groups and hence the only possible cause for an observed effect). As such, “RCTs trust to procedure [while] other methods import information” to control for unknown confounders. Accordingly, it is claimed that RCTs have the highest internal validity of all trial methodologies and therefore RCTs should be regarded as the gold standard.

The claims of EBM proponents sound good in theory. However, as John Worrall has convincingly pointed out, randomization does not, in practice, “control all at once for all possible confounders – known and unknown; and hence at a stroke both do away with any reliance on judgment about background knowledge and eliminate the worry about incompleteness.” He insists: “no one really believes [this].” And he is correct. In practice researchers acknowledge that randomization can sometimes imperfectly distribute confounders, which is why, before

---

182 Ibid., 359.
conducting a trial, researchers look for baseline imbalances (known confounders) between the control and experimental groups. This sounds very similar to what researchers conducting observational studies do, although rather than double-checking the results of randomization they attempt to distribute confounders by hand. The real issue then is whether randomization actually controls for confounders to the extent EBM proponents claim it does, since it appears that even known confounding factors fail to be controlled for at all times. The problem, as Worrall points out, “is that randomization can control only for most, but not for all, possible confounding factors…when there are indefinitely many factors, both known and unknown, which may lead to bias.”183 Consequently, given the existence of infinitely many confounders (both known and unknown) it is likely that at least one (if not more) confounder will not be controlled for leading to the introduction of bias. Consequently, the control group will be different from the experimental group in some, possibly significant, way. As such, the confounder may produce a positive (or negative) effect in the experimental group, which would then be attributed to the experimental treatment rather than the actual cause of the effect (i.e., the uncontrolled for confounder). That is, the confounder would “provide an alternative hypothesis for the result.”184

How does one resolve this problem? Given infinitely many confounders, one must re-randomize the subjects and repeat the study multiple times and then average (perform a meta-analysis on) the data collected over time. This is the only way one can be sure that the causal link between the demonstrated treatment effect and the experimental treatment holds. Given that it is infeasible in practice to replicate a study several times, it appears that RCTs cannot in practice

183 Solomon, Making Medical Knowledge, 106.
184 Howick, The Philosophy of Evidence-Based Medicine, 56.
claim a “special epistemic power” for randomization, making its use as a selection criterion for the evidence hierarchy epistemically indefensible.

Miriam Solomon finds this conclusion problematic as it “only considers the abstract possibility of multiple unknown variables…and does not consider the likely relationship (correlation) of those variables with one another.” That is, she finds it unlikely that in practice there are unknown confounders, uncontrolled for, that actually interfere with the causal relations between treatments and effects. She also points to “the common requirement for at least one successful replication” of a trial to gain FDA approval for an intervention. A successful replication is meant to serve as evidence that all confounders have been controlled for and no variable has interfered with the observation of treatment effects. However, as Solomon herself later points out, replications of studies are disappointingly unsuccessful in more than half of attempts. She draws on a study conducted by Ioannidis in 2005, which evaluated the attempted replications of 59 highly cited original studies. Citing Ioannidis’ results she reports, “Less than half (44 percent) were replicated with the same results; 16 percent of [sic] were contradicted by attempted replications and 16 percent of the replications found the effect to be smaller than the original study; the rest (24 percent) were neither repeated nor challenged.” It is possible, maybe even probable, that the replications failed for reasons other than uncontrolled for confounding factors, however there is no way to know for sure. What we do know for sure is that, in the words of Kirstin Borgerson, “defenders of the special causal ability of RCTs make claims about the epistemic powers of actual RCTs based on what would happen in ideal

---

186 Solomon, Making Medical Knowledge, 136.
187 Ibid., 136.
188 Ibid., 148.
RCTs.”189 Borgerson’s point is well taken. We have reason to be highly confident in the internal validity of ideal RCTs, however we do not know exactly how well real RCTs secure causal claims. Given the prevalence of conflicting evidence from RCTs it seems we are at least somewhat justified in doubting the success of randomization in ruling out all confounders. Accordingly, it is inappropriate to use the ability of ideal RCTs to clinch causal claims as a yardstick for all other methodologies.

III.II Minimizing Bias

The second justification proponents of the hierarchy frequently give is that RCTs produce less biased results than any other methodology. According to Richard Ashcroft, the evidence hierarchy is based on the assumption that “it is possible to rank methods of inquiry by their susceptibility to bias.”190 In this case the particular biases being considered are ascertainment bias and selection bias. Ascertainment bias is present when the group assignments of individual subjects are known and, consequently, one is biased (consciously or subconsciously) in their assessments of the trial. For example, a subject may figure out that they are receiving the experimental treatment in a study either by accident or due to side effects of the treatment. Consequently, they may report more positive effects to make the investigator happy. Alternatively, an investigator may determine that a subject is in the control group due to a lack of side effects and, as a consequence, spend less time talking to the patient and assessing her medical state. In both these cases, knowledge of the subject’s group assignment biases the data reported and/or collected.

Selection bias is present when subjects have not been properly randomized to the study groups and consequently differ in baseline characteristics (e.g., co-morbidities, age, etc.). An example would be if a clinician, knowing the randomization algorithm of the study, was able to anticipate that the next subject would be assigned to the intervention group and decided not to enroll his 60 year old patient with multiple morbidities in the hopes that a younger, healthier patient would take her spot, increasing the likelihood that a significant positive effect for the intervention group would be identified. It may also be the case that a clinician would postpone enrolling a very ill patient who wasn’t responding to any other treatments if he knew the patient would be assigned to the control group since, if assigned to the intervention group, the patient might respond positively.

In the case of ascertainment bias, the primary way in which it can be prevented is by concealing the group assignment of subjects from the subject, the investigator, the medical staff involved in the care of subjects, and even those involved in the statistical analyses of trial data. It is important to note that this cannot be achieved by randomizing, as randomization only effects the allocation of subjects into either the intervention group or the control group. The key methodological feature to prevent ascertainment bias is masking. Almost all RCTs employ masking to prevent ascertainment bias, however other methodologies, such as cohort studies, also employ masking and therefore are equally capable of controlling for ascertainment bias. Additionally, for some methodologies, such as case studies, which are analyses of individual patients, ascertainment bias is not a factor because there is no group assignment to conceal. Given that ascertainment bias can be equally well controlled for through the use of masking in other study methodologies and isn’t even an issue in other methodologies, the RCT cannot claim a superior epistemic status on the grounds of preventing ascertainment bias.
Returning to selection bias, the primary concern here is with avoiding “systematic error in creating intervention groups.”\(^{191}\) However, even though the allocation of subjects into the control and intervention groups is determined through the process of randomization, randomization cannot prevent selection bias. Selection bias can only be prevented through “some form of allocation concealment,”\(^{192}\) that is, by suppressing the algorithm used to allocate patients to different study groups. Whether allocation is determined by a computer algorithm or something as simple as the month one was born in (i.e., January through June in the control group and July through December in the intervention group), the important factor is that investigators and subjects remain unaware of the algorithm.

It should be mentioned that selection bias is not only an issue for randomized studies. Cohort studies also employ allocation principles that, while not randomized, should be concealed from the investigators and patients, which can just as easily be achieved through allocation concealment. Further, other types of studies, like qualitative studies, do not allocate subjects into groups and therefore selection bias does not pose a problem. Much like ascertainment bias, a methodological feature unique to RCTs does not resolve selection bias. Other methodologies are equally able to control for selection bias and for others it’s not even an issue. Once again, it appears RCTs cannot claim special epistemic status on the grounds of better controlling for bias. In the words of Kirstin Borgerson, “The one or two biases that RCTs allegedly eliminate are either equally well managed by other methods (because they are not necessarily connected to randomization), or they are not necessarily encountered by other methods. As such, the claim

\(^{192}\) Ibid., 227.
that RCTs, by design, necessarily produce less biased results than other trial methodologies is false.**193**

At this point it has been established that the two primary justifications set forth in defense of the evidence hierarchy and the gold standard status of RCTs fail. Without any additional epistemic justifications for why the RCT methodology should be considered superior to all others, it appears that the evidence hierarchy, and the rules of evidence underpinning it, amounts to nothing more than a “cultural agreement to privilege RCTs.”**194**

**IV. Further Critique: External Validity**

The previous section identified and addressed those issues concerning the internal validity of RCTs, that is, determining whether they warrant causal conclusions and to what extent they can successfully minimize bias. This section focuses on issues regarding the external validity of RCTs. If one remains unconvinced by the arguments advanced above, one must consider the significant practical problems clinicians face in practice when trying to externalize results from RCTs. Richard Horton has claimed “issues of external validity are the most important that face clinical research today and that the failure to resolve them is largely responsible for the indifference doctors worldwide show towards research evidence.”**195** Rothwell strengthens this assertion with his claim that the “lack of consideration of external validity is the most frequent criticism by clinicians of RCTs, systematic reviews, and guidelines.”**196** The particular concern is whether the results we get from a trial can be applied to individual patients in routine practice. Peter Rothwell has defined external validity as “whether the results [from an RCT] can be

---

193 Ibid., 227.
196 Rothwell, “External Validity,” 82.
reasonably applied to a definable group of patients in a particular clinical setting in routine practice.”\textsuperscript{197} The ability of a trial to successfully export its results to patients in the clinic has come to be known in the literature as “externalization” or “generalizability.” While EBM’s proponents have not been so bold as to assert that RCTs externalize well, they have largely ignored the issue, leading, as Horton pointed out, to clinicians largely ignoring the information provided by RCTs.

To give a more precise explanation of externalization and the moves required to achieve it, it is necessary to start by parsing out the causal issues at play. Nancy Cartwright gives an exceptionally clear account of how one moves from the causal claims established in an ideal RCT to an effectiveness claim for a target population (i.e., the patient population of the particular patient one is concerned with treating). Cartwright’s process can be represented in three steps:

1. “RCTs establish causal claims...Given the probabilistic theory of causality it follows formally that positive results in an ideal RCT with treatment C and outcome E deductively implies ‘C causes E in the experimental population’.”\textsuperscript{198}

2. We then “move to predict what would happen in a new RCT in the target population [the population we want to treat], or a subpopulation of it, from what happens in an RCT in a different population [the experimental population]. If we follow standard usage and describe RCT results as efficacy results, the inference here is roughly from efficacy in the experimental population to efficacy in the target. [Efficacy is the ability to produce a desired effect in ideal circumstances.] [This does not establish] that C will actually result in E when implemented there [in the target population].”\textsuperscript{199}

3. We “move from ‘C causes E in the target population’ to ‘C will result in E if implemented in this or that way’.\textsuperscript{200} That is, we move from an efficacy claim to an effectiveness claim (i.e., the ability to produce an effect in actual usage) in the target population.

What Cartwright is effectively describing is the need to apply the results provided for an “abstract average patient” in the initial trial to a second abstract patient representing the target

\textsuperscript{197} Ibid., 82
\textsuperscript{198} N. Cartwright, “What Are Randomised Controlled Trials Good For?” \textit{Philosophical Studies} 147 (2010): 68.
\textsuperscript{199} Ibid., 68-69.
\textsuperscript{200} Ibid., 69.
population. Then, in order to establish an effectiveness claim, the results from the second
“abstract patient” would have to be particularized to the patient being treated in the clinic. It
should be noted that the reason RCTs can only provide information for an “abstract patient” is
because, as Tonelli describes it, “The randomized controlled trial must obscure the effect of
individual variability, resulting in knowledge that is referable only to an “average” patient.” 201
This is primarily due to the methodological features of the RCT meant to ensure its ability to
identify causal relationships.

In theory, the process of externalization Cartwright is describing appears simple.
However, in practice it is much more complicated. Given how causal claims are achieved by the
RCT’s methodology, there is a lot of important information we would need to know to be able to
export the claim ‘C causes E’ to another population. In particular, we would need to know that
the causal structures of the two populations were similar in all relevant ways. That is, any
confounding factors affecting the causal relationship between C and E would need to be similarly
controlled for in each case. Given that there are many unknown confounders, it would be hard to
determine whether the causal structures between the two populations were in fact the same. As
Cartwright describes it,

Two populations share the ‘same causal structure with respect to the causal
principle ‘C causes E ’ from the point of view of the probabilistic theory of
causality iff the two populations share the same reasons for dependencies to
appear or disappear between C and E (i.e. the same choice of factors from which
to form the state descriptions Ki ) and the same conditional probabilities of E
given C in each. 202

Given that it is difficult to access complete knowledge of the causal structure of each population
it is very difficult to determine whether ‘C causes E’ will hold in both cases. Therefore, one can
only make educated judgments about whether a causal relationship (i.e., a treatment effect) will

202 Cartwright, “What Are Randomised Controlled Trials Good For?” 66.
in fact hold between different populations. There is no way to know for sure. As Solomon states, “the result is, at best, confidence (not certainty) that an intervention will work in a specific new setting.”

The third step in Cartwright’s process involves turning an efficacy claim into an effectiveness claim. In the first two steps Cartwright is addressing how to move from the RCT’s study population to a target population, more closely resembling the patients one will treat in their practice. In the third step, one needs to move from the target population to the individual patient. Cartwright in her article is concerned with externalizing causal claims for public policy rather than individual patients so the discussion must depart from hers at this point.

When dealing with patients in the clinical setting there are a host of issues that will effect how well efficacy claims will translate into effectiveness claims. As Jonathan Fuller states, “trials map poorly onto the reality of clinical medicine; the trial participants are not representative of patients routinely encountered by clinicians, and the ideal and pristine trial conditions are too dissimilar from the concrete and messy world of clinical practice.” Accordingly, a number of new concerns are introduced. Trial conditions and populations differ from those “in the wild,” as John Worrall likes to say, in four primary ways: 1) patient attributes, 2) treatment details, 3) setting, and 4) timetable. However, patient attributes and treatment details will be the only two discussed in this section as they are sufficient to demonstrate the difficulty RCTs have in externalizing their results.

---


Patient Attributes

As mentioned above, ideal RCTs secure efficacy claims for abstract patients. However, it is also the case that these abstract patients are in many ways also “ideal patients.” That is, due to a number of features of how RCTs are conducted many subjects are screened out or prevented from participating in trials because they have average characteristics such as multimorbidities or old age. Some of the patient attributes that can influence the externalizability of RCT results include, “age (effectiveness in younger or older patients), gender, (effectiveness generally), severity of the disease (effectiveness in milder or [more severe] forms of the condition), risk factors (effectiveness in patients with risk factors for the condition (e.g., smokers)), co-morbidities (influence of other conditions on effectiveness), ethnicity (effectiveness in other ethnic groups), and socioeconomic status (effectiveness in disadvantaged patients).” All of these factors have been known to affect the effectiveness of some interventions for particular patient populations. This does not mean these attributes will always be salient factors when one externalizes trial data, but they are attributes clinicians need to consider and be aware of. This is a primary example of clinical experience and pathophysiological rationale being integral to the practice of medicine because both are required to determine whether any of these attributes may have salient effects and if so, in what ways.

In practice, many of these attributes are intentionally screened out of RCT study populations. In some instances they are screened out intentionally because they are known confounders, such as comorbidities, which would make establishing an efficacy claim much more difficult. Other attributes may inadvertently be excluded from RCT study populations for a number of reasons (e.g., ethnic homogeneity in the geographic area in which the trial is being conducted).

206 M. Rawlins, “De Testimonio: On the Evidence for Decisions about the Use of Therapeutic Interventions,” Clinical Medicine 8(2008): 582. The list given has been adapted from a table provide by Rawlins in this essay.
performed). Peter Rothwell provides an exceptional list of the ways patients are systematically excluded from trials on the basis of these attributes. In his list he includes: eligibility criteria, exclusion criteria, and run-in periods (i.e., pre-trial periods where subjects are screened out if they cannot adhere to the treatment regiment or if they demonstrate a placebo response). The sheer number of attributes capable of interfering with externalization, in addition to the inadvertent and intentional ways patients with these attributes are systematically excluded from trials, indicates a significant problem for RCTs. Accordingly, for these reasons, it is understandable that clinicians frequently do not consult the research literature.

If one doubts the frequency with which these attributes are excluded from studies the following statistics provided by Fuller are illuminating:

A systematic review of 283 RCTs published in major medical journals found that 81.3% of trials excluded patients due to common co-morbidities, 38.5% due to older age and 54.1% due to concurrent use of commonly prescribed medications. Medical co-morbidities and concurrent medication use served as exclusion criteria even more often among the subset of drug intervention trials. Yet, most patients with at least one chronic disease have multiple, and we often use multiple medications to manage multiple chronic diseases.

From these statistics, it is clear that the exclusion of these attributes from study populations is not insignificant. Additionally, as Fuller points out, in practice most of the patients who would be receiving drug interventions (i.e., those with co-morbidities who use multiple medications) are screened out of the trials that will ultimately be used (in theory) to inform their care. In this case especially, the tension between maintaining the internal validity of a trial and the need to externalize the trials results is significant and has real consequences for real patients. The value of the research literature certainly becomes questionable if it cannot be used to inform the care of

---


209 Fuller, “Rationality and the Generalization,” 644.
the patients who need the interventions most.

Treatment Details

The particulars of how treatments are administered in the research setting, in comparison to the clinical setting, also introduce variables that can affect the effectiveness of a treatment in one’s target population. Accordingly, the original study’s efficacy claims do not track. There are five primary differences between the research and clinical setting: 1) dosing (too high a dose is often used in RCTs), 2) timing of administration (compliance, i.e., patients are closely monitored in trials to ensure they adhere to the treatment protocol – this does not happen in practice), 3) duration of therapy (RCTs often test interventions that are typically taken long-term, e.g., blood pressure medications, for short time periods), 4) co-medications (RCTs typically prevent patients from taking many medications limiting the available information regarding adverse interactions), and 5) comparative effectiveness (many trials test interventions against placebo which does not indicate effectiveness in comparison to other products).210 As one can see from the types of variables introduced between the research and clinical setting most are the consequence of particular trial designs. The issues here are driven by concerns over preserving the internal validity of the RCT (i.e., 2, 4, and 5), ensuring significant positive outcomes (i.e., 1), and most likely the limitations of resources (i.e., 3).

As was the case with patient variables, treatment variables require clinical expertise and pathophysiologic rationale to determine whether an intervention will be effective in the target population. However, while clinical expertise and pathophysiologic rationale go some distance in identifying the salient variables that differ between the treatment population and the target population...
population, at the most they can be considered well-educated guesses about how the intervention will interact given the known particularities of the patient population. In some cases this is an impossible task that leads clinicians to ignore the recommendations from the research literature and/or clinical guidelines. Accordingly, the problem of external validity for RCTs is significant and has most likely been one (of many) contributing factors to clinicians’ resistance to EBM.

Given the complex and varied ways in which different patient attributes and trial protocols for dispensing treatment can effect the externalizability of trial results for relevant patient populations, it appears there is a significant tension preventing RCTs from maintaining both their internal and external validity. The concerns of internal validity that require such tight controls over variables make the causal claims they indentify applicable to a very narrow population, most of the time not including the patient population in most need of the intervention. Some suggestions have been made for addressing external validity, such as relying on inclusion and exclusion criteria as applicability guidelines in addition to looking for compelling reasons why the trial should not apply (i.e., applying clinical expertise and pathophysiological rationale). While such an approach is prudent, it still does not solve the primary problem of externalizability and, in some cases, can lead to disastrous results.

Given the difficulties clinicians have in externalizing trial results to their patients, it stands to reason that the evidence hierarchy, in considering the epistemic superiority of different methodologies, should include the clinical usefulness of the evidence produced as one of it ranking criterion. Otherwise, the hierarchy can be considered nothing more than an ideal ranking schema unfit for application in an applied science like medicine.

---

211 Fuller, “Rationality and the Generalization, 645.
V. Conclusion

In this chapter the two primary epistemic justifications for the RCT’s gold standard status, and the consequent ranking schema of the evidence hierarchy, have been challenged. The purported superior ability of RCTs to identify causal claims has been disproven by demonstrating that such causal claims can only be identified in ideal RCTs. Additionally, the claim that RCTs produce less biased results has also been disproven by demonstrating that randomization, the only methodological feature unique to RCTs, does not in fact prevent the biases it is claimed to. Furthermore, other methodologies, such as observational studies, are equally able to control for the same biases. Lastly, an examination of the difficulties RCTs have in externalizing their results was undertaken. It was demonstrated that much of the evidence produced by RCTs is difficult, if not impossible in some cases, to apply in practice. Consequently, it has been suggested that “usefulness of information” should be added to the evaluative criterion used to assess methodologies on the evidence hierarchy. Taking all of these conclusions into consideration, it appears there is not epistemic justification available to RCTs to defend their gold standard status.
BIBLIOGRAPHY


Rothwell, P.M. “External Validity of Randomised Controlled Trials: “To Whom Do the Results of this Trial Apply?”” *The Lancet* 365 (2005): 82-93.


APPENDIX A:
MECHANISTIC EVIDENCE IN EVIDENCE-BASED MEDICINE:
ANALYZING THE RUSSO-WILLIAMSON THESIS

I. Introduction

In 1992 evidence-based medicine (EBM) became the dominant paradigm for the practice of medicine. One of the primary factors influencing this shift was the growing distrust of "pathophysiological rationale" or, in other words, the use of mechanistic reasoning for making treatment decisions. Part of the justification for this shift recognizes that mechanistic reasoning in medicine has tended to be less reliable than mechanistic reasoning in the basic sciences because the relationship between explanation and prediction come apart due to the complexity of the causal environment in the human body as well as the unpredictability of well-understood interactions between mechanisms when they are under intervention. These factors explain why mechanistic evidence can be somewhat unreliable for predicting the efficacy of an intervention for a patient population. As a consequence, EBM has systematically devalued and discouraged the use of mechanistic evidence in clinical decision-making and has prioritized the use of difference-making evidence generated by tightly controlled comparative clinical studies. These values are reflected in EBM's evidence hierarchy with randomized controlled trials as the gold standard at the top and mechanistic reasoning either absent or at the bottom.

In recent years a number of philosophers have sought to reintroduce mechanistic reasoning to the practice of evidence-based medicine and causal assessment in the health sciences. The importance and role carved out for mechanistic evidence has varied widely

---

between the various accounts, however, common among them is the claim that mechanistic evidence deserves recognition as a legitimate source of evidence within the theory and practice of EBM. The most demanding and controversial of these accounts is the Russo-Williamson Thesis, which requires (1) the acquisition of both mechanist and difference-making evidence to establish a causal claim and (2) that both types of evidence be treated with equal importance in causal assessment.

Given the controversy surrounding different interpretations of the RWT, this project will seek to determine and assess the most plausible instantiation of the RWT. The analysis will begin by assessing Russo and Williamson’s (R&W) original presentations of the RWT, which will ultimately be deemed imprecise and at times self-contradictory. Consequently, alternate interpretations of the RWT put forth by Russo, Gillies, and Illari will be considered. It will then be argued that the most reasonable interpretation of the RWT is Illari’s weak interpretation, which will be referred to as the VWRWT (Very Weak Russo-Williamson Thesis).

---


215 This project focuses solely on Russo and Williamson’s account of the RWT in their 2007 and 2011 papers. It should be recognized that Russo and Williamson have published additional papers since these original essays further elaborating their account. Acknowledging this, this project seeks to address the original formulation of the RWT and the consequent conversation that occurred in the literature because there is something important and salvageable in the RWT even given its most uncharitable reading. This project seeks to emphasize that point.


Once the VWRWT has been established, concerns over whether this view is trivial or too far removed from the original theory will be considered. Lastly, the VWRWT will be evaluated in light of Jeremy Howick’s two desiderata for mechanistic evidence, which will result in a revised final version of the VWRWT. In conclusion, it will be held that the VWRWT is a reasonable and plausible interpretation of the RWT and, therefore, barring further objections, should be taken seriously by practitioners of EBM.

II. The Russo-Williamson Thesis

The strongest version of the RWT is found in R&W’s 2007 paper, “Interpreting Causality in the Health Sciences,” which states, “in the health sciences…both mechanistic and probabilistic evidence are required to substantiate causal claims.”220 Reasonably, philosophers have interpreted this instantiation of the RWT as holding that “mechanistic reasoning is required alongside comparative clinical studies”221 when attempting to establish causal claims in medicine, in particular, the efficacy of interventions. Such a strong requirement, however, can have counter-productive results, which is evidenced by ample historical examples. Take for example the Semmelweis case. According to the RWT, Semmelweis’ contemporaries were justified in dismissing his probabilistic evidence that a lack of hand washing by doctors was responsible for the high death rate from puerperal fever of the birthing mothers they assisted because there was no clear mechanism grounding his theory.222 Also consider a more contemporary example. According to the RWT, to establish the efficacy of the mother’s kiss

221 Howick, The Philosophy of Evidence-Based Medicine, 135.
222 It should be noted that proper judgment of the Semmelweis case in this context is controversial. Russo and Williamson argue that Semmelweis’ contemporaries were right to reject his theory until germ theory was established. However, the example here is being used to probe the opposite intuition, i.e., that it was unreasonable for Semmelweis’ contemporaries to reject his proposal to institute hand washing due to a lack of mechanistic evidence.
technique for dislodging a marble from a child’s nose (i.e., obstructing one nostril while blowing into the other) probabilistic evidence would be required in addition to the observed simple mechanism. In these two examples, the negative consequences of the strong view are evident. In the Semmelweis case, women continued to die because a mechanism could not be found to support what Semmelweis already had overwhelming statistical evidence of. In the mother’s kiss case, a demand for difference-making evidence would not only be a waste of resources, it may also be unethical depending on what studies one chose to conduct. One can generate many more counter examples to the strong RWT making it clear that this interpretation of the RWT is simply untenable in practice. What should be clear from these cases is that historical and contemporary examples demonstrate that not all causal claims require both kinds of evidence and insistence on the fulfillment of such a requirement can be dangerous, wasteful, and/or unnecessary. It even appears that R&W accept this fact in their 2011 essay in which they state, “the biomedical sciences contain compelling cases in which it is clear that neither good evidence of difference-making nor good evidence of mechanisms is enough on its own to establish a causal claim; typically one needs both.” The reference to only “compelling cases” and the use of “typically” indicates that R&W have recognized the limitations of the strongest interpretation of their view, i.e., not all cases can or should require mixed evidence.

Given the implausibility of the strong view as constructed above and R&W’s partial abandonment of it, one might wonder what position they are in fact advocating for. In their 2011 essay, R&W make it clear that the primary goal of their account is to provide an alternative theory of causation in the context of disease: the epistemic theory. Their motivation for introducing such a theory was made explicit in their 2007 essay: “unfortunately, both types of

---

224 Howick, The Philosophy of Evidence-Based Medicine, 154-157.
[causal] account fail to handle the dual aspect of causal epistemology. Neither the mechanistic or the difference-making approach can adequately account for two types of evidence—mechanistic and probabilistic—for a single causal claim. In 2011 they reinforce this motivation for their account, stating:

Only under the epistemic theory is it possible, let alone plausible, to include mechanistic evidence alongside difference-making evidence in a completely egalitarian way. The reason is that the epistemic theory is the only account (i) that does not reduce the meaning of causation to a single evidential component (i.e., to mechanisms or to difference-making) and (ii) that treats these two kinds of evidence on a par in causal assessment.

Here they are clearly arguing that on mechanistic and difference-making theories of causation only one type of evidence can be accounted for and therefore only one type of evidence can factor into causal assessments. R&W rightly find this problematic because “in the health sciences it is… commonplace that both mechanistic and probabilistic evidence are required to substantiate causal claims.” If one takes this to be a general statement about causal claims, and not about each individual causal claim, one can see that R&W are sensitive to the fact that health professionals rely on both types of evidence to make causal assessments and therefore need the epistemic theory. The epistemic theory, however, does not necessarily entail that each type of evidence is required for each and every causal claim established. It does, however, mean that any theory of causation that cannot account for mixed evidence is not a viable theory of causation in the context of disease. Given that health professionals do “infer causal relations from mixed evidence,” R&W’s epistemic theory appears reasonable, although the strong evidentiary requirements they draw from it are not and require revision.

---

229 Ibid., 158.
III. Interpretations of the RWT

In this section, weak interpretations of the RWT by Gillies and Illari—who are skeptical of the strong RWT but are also generous in their treatment of it—will be examined. However, before considering these interpretations it will be productive to consider another essay written by Frederica Russo—in which she discusses the relationship between mechanistic and difference-making evidence—and what implications her claims may have for the RWT.230

III.I. The Russo-Based Interpretation

In her essay, “Causal Webs in Epidemiology,” Russo discusses mechanistic and difference-making evidence and the relationship between the two. Her discussions on this issue reveal the much broader concept of mechanistic evidence with which she is working that never quite gets discussed in her 2007 and 2011 works with Williamson. Without access to her full understanding of what mechanistic evidence is or may be, one might assume that when requiring “mechanistic evidence” to establish a causal claim in the RWT, it refers to a complete description of a confirmed linking mechanism from which inferences can be made about treatment effects. However, in Russo’s 2011 essay, she considers “plausible mechanism[s] to be tested and (dis)confirmed afterwards”231 and “any information that refers to the functioning or modus operandi underlying the phenomenon under investigation”232 as legitimate mechanistic evidence. If this is the definition she intended for use in the RWT, then the strong interpretation doesn’t seem as daunting because it is fairly easy to satisfy. Most randomized controlled trials and other comparative studies invoke some type of plausible mechanism and/or some knowledge of how the relevant intervention works in its trial design and justification. One may reasonably object at

---

231 Ibid., 84
232 Ibid., 84
this point that anyone can generate some type of mechanism or *modus operandi* to explain the causal link of an intervention, effectively making the requirement meaningless. This objection is accurate and will be revisited in a slightly stronger form;²³³ however, another objection stemming from Russo’s further comments needs to be addressed.

In her essay, Russo also claims that mechanistic and difference-making evidence are not necessarily “disentangled” from each other. She states, “mechanistic information is most often used *implicitly* in randomised clinical trials (RCTs)…they… use information about known or plausible mechanisms, their parts or functioning, when they set up and evaluate the trials.”²³⁴ If Russo is willing to satisfy the mechanistic evidence requirement with the mechanistic evidence *implicit* in RCTs, then it appears that for those counterexamples where critics claim only difference-making evidence is required, she should not consider them real counterexamples because both types of evidence are in fact present. In this case, there also appears to be no theoretical or strategic advantage to having the strong RWT requirement for causal assessment: the requirement is trivial if either type of evidence implicitly contains the other and the unnecessary theoretical and interpretative complexity it causes cannot be considered a virtue. It is highly unlikely that Russo would consider this type of mechanistic evidence sufficient to satisfy the requirement of the RWT, however, the objection still needs to be made.

Lastly, Russo appears to all but explicitly dismiss the strong RWT requirement when she states,

[It does not give a] checklist for causal relations: if mechanistic and difference-making evidence are both present, infer causation, and don’t otherwise. Mechanistic and difference-making evidence are not monolithic blocks that are either present or absent. Causal analysis is far from being a schematic and algorithmic activity with step-wise procedures and checklists to tick. Understanding what kind of evidence is needed in order

²³³ See the objection waged against Gillies in this essay.
²³⁴ Russo, “Causal Webs,” 84.
to establish causal relations is tantamount to understanding what is the relevant information needed in order to draw sensible and reliable conclusions from the data available.\textsuperscript{235}

In this passage Russo is setting forth a position she refers to as “evidential pluralism.” She explicitly mentions that this discussion is an extension and refinement of the view she put forth with Williamson in 2007. Considered as such, Russo has effectively ruled out the strong interpretation of the RWT by stating that the requirement is not hard and fast and that the constraints on causal analysis really should be relative to the causal relations in question. If this weaker and more reasonable approach is an accurate interpretation of the RWT, it is a much more plausible position. However, mention of this nuanced approach to the RWT remains wholly absent from R&W’s 2011 articulation of their position, aside from the minor concessions mentioned above. Consequently it remains unclear what motivation R&W have not to expound this more reasonable position.

Given the residual unclarity regarding how to interpret the RWT even after investigating further Russo’s position, one cannot claim for certain that R&W do not endorse the strong interpretation. Too many details are missing. Russo’s revised view either abandons the essential requirement of the RWT (i.e., that both mechanistic and difference-making evidence be required for causal analysis) or it makes the requirement trivial. Beyond this, the question still remains as to what a weaker interpretation would entail. As such, it is worthwhile to investigate whether a weaker interpretation of the RWT, which preserves its essential requirements, can be articulated. In the sections that follow, interpretations set forth by Gillies\textsuperscript{236} and Illari\textsuperscript{237} will be considered and assessed.

\textsuperscript{235} Ibid., 80.
\textsuperscript{237} Illari, “Mechanistic Evidence,” 139-157.
III.II Gillies’ Interpretation

Donald Gillies, in his 2011 essay, “The Russo-Williamson Thesis and the Question of Whether Smoking Causes Heart Disease,” articulates strong and weak interpretations of the RWT. The difference between his interpretations hinges on a distinction between a plausible mechanism (weak RWT) and a confirmed mechanism (strong RWT). He states,

“In order to establish that A causes B, observational statistical evidence does not suffice. Such evidence needs to be supplemented by interventional evidence, which can take the form of showing that there is a plausible mechanism linking A to B,” where “A plausible mechanism is one confirmed by our general background knowledge but not necessarily by particular investigations and experiments designed to test it out.”

This interpretation employs a concept of plausible mechanism similar to Russo’s. Gillies’ account is a little stronger because he adds in the constraint that it must be confirmed by our general background knowledge. In this context, “general background knowledge” can be understood as the current scientific knowledge relevant to the case at hand. A Kuhnian might describe it as the information provided by the current scientific paradigm. As such, even though this addition grounds the mechanism in scientific knowledge, requiring this type of confirmation is also problematic. Having the plausibility of a mechanism hinge on its accordance with the current scientific paradigm threatens to stagnate medical advancements. By requiring that each causal claim be supported by either a known or plausible mechanism, an investigator who has overwhelming difference-making evidence of the efficacy of a breakthrough intervention whose mechanism is not known will be constrained if the current medical/scientific paradigm is unable to conceptualize its causal mechanism.

239 Alex Broadbent advances a similar yet slightly different objection when he states, “the mechanistic stance has on occasion led to a bias towards existing knowledge” (Broadbent, “Inferring Causation,” 61). Here Broadbent is referring to the idea that the mere search for a mechanism tends to direct us towards our current knowledge because that is where we are most likely to find the resources for generating a plausible or confirmed (Broadbent doesn’t distinguish) mechanism. Broadbent’s objection is more general than the one waged against Gillies account and,
fair much better than the original RWT because it cannot apply to all causal relations and, if enforced, may hinder scientific progress in select cases.

III.III. Illari’s Interpretation & The VWRWT

Phyllis McKay Illari does an excellent job analyzing and setting forth various interpretations of the RWT. Her weak account maintains that “there are two kinds of things [objects] we have evidence of: what you have evidence of when you have evidence of a mechanism, and when you have evidence of difference-making.”

This account, as Illari has articulated it, focuses on the object of evidence and the conclusions that can be drawn from them: “evidence of difference-making is just evidence that the effect does indeed vary with the postulated cause” and “evidence of mechanism is just evidence of the existence of a mechanism or mechanisms in the domain of inquiry.”

Illari understands the motivation for requiring both types of evidence on the weak interpretation to be related to the known problems with each type of evidence—difference-making evidence suffers from confounding and mechanistic evidence suffers from masking. Recognizing that each weakness of one type of evidence is the strength of the other, Illari states, “Russo and Williamson are right to draw attention to the importance of mechanisms in causal inference. This is due to the different roles evidence of mechanisms and of difference-making play in causal inference.”

R&W (2007) briefly discuss the benefit of having both kinds of evidence, but do so primarily from the perspective of making it clear that mechanisms prevent spurious causal claims supported by difference-making evidence: “mechanisms…impose

\[\text{[240]}\]

Illari, “Mechanistic Evidence,” 141.

\[\text{241}\]

Ibid., 144.

\[\text{242}\]

Ibid., 145.

\[\text{243}\]

Ibid., 148.
negative constraints: if there is no plausible mechanism from $C$ to $E$, then any correlation is likely to be spurious.”\textsuperscript{244} However, once again, one should notice that as R&W have stated the requirement, it implies that any causal relation that is not supported by a mechanism will be spurious. Remembering the Semmelweis example, there are clearly cases in which this is false and, therefore, the requirement is too strong and undesirable. Consequently, even though it may be the case that the combination of both types of evidence will tend to be more effective at ruling out confounding and masking, R&W have once again overstated the usefulness of mixed evidence, causing their approach to demand too much.\textsuperscript{245}

In an effort to weaken the RWT requirement that both types of evidence be present, Illari considers which of the various potential forms of mechanistic evidence fulfill the requirement. She discusses four types of mechanistic evidence:

1. Evidence of *what the mechanism is* in detail.
2. Evidence *that there is* a mechanism of the postulated kind [incomplete mechanism].
3. Postulated mechanism [plausible mechanism], based on evidence of analogous mechanisms.
4. Evidence that there is *no* mechanism.\textsuperscript{246}

She categorizes these four types further: types 1 and 2 are evidence of a linking mechanism, type 3 is evidence from other mechanisms in the domain, and type 4 is evidence from mechanisms in the same or other domains. Illari ultimately argues that requiring mechanistic evidence of type 1

\textsuperscript{244} Russo and Williamson, “Interpreting Causality,” 159.
\textsuperscript{245} Jeremy Howick suggests that a reasonable weak interpretation of the RWT would be “to alter the claim that mechanistic reasoning is required to rule out spurious relationships and argue instead that hypotheses supported by both comparative clinical studies and mechanistic reasoning are less likely to be spurious than hypotheses supported by one type of evidence alone” (Howick, *The Philosophy of Evidence-Based Medicine*, 136). Howick’s modified version is more reasonable because it acknowledges that mechanistic reasoning is not always required to rule out spuriousness, while also recognizing the strength of having both types of evidence. However, as Howick observes, this weaker version “can no longer be interpreted as [R&W’s] thesis” (Ibid., 136) because it directly contradicts the original statement of the RWT and fails to satisfy its first requirement.
\textsuperscript{246} Illari, “Mechanistic Evidence,” 151.
or of type 2 is too strong for reasons similar to those discussed in other sections. However, she defends the RWT by requiring mechanistic evidence of type 3 or 4.

Type 3 evidence boils down to what has previously been referred to as a plausible mechanism. Because plausible mechanisms are constrained by what our current medical/scientific paradigm is capable of comprehending, without any verification (partial or complete) of the mechanism being considered, this requirement remains too strong. To clarify, the objection is not to the usefulness of plausible mechanisms in indicating the existence of a mechanism; it is to the requirement that every intervention must demonstrate, at the least, that there is a mechanism supported by background knowledge. In those cases where the mechanism defies current scientific knowledge, this requirement must be considered too strong.

Type 4 evidence “allows the possibility of positive evidence of absence” and, when formulated for the RWT, it requires “no positive evidence of the absence of a mechanism linking C and E.” This type of evidence can come from other disciplines, such as physics or chemistry, what we know about other mechanisms in the domain, or acknowledged constraints on the domain one is dealing with that would prevent the type of mechanism in question. This type of evidence appears to be the only type that could reasonably be required to establish a causal claim in every case. It stands to reason that a causal claim cannot readily be established if there is positive evidence indicating the absence of a mechanism. In such cases, such evidence strongly indicates that any difference-making evidence in support of a causal claim is the result of confounding and, therefore, further investigation is required.

It appears at this point that an acceptable albeit very weak version of the RWT (VWRWT) can be articulated, which states

247 See discussion of Gillies’ plausible mechanisms in section III.II.
248 Illari, “Mechanistic Evidence,” 153,
249 Ibid., 155.
Both mechanistic and probabilistic evidence are required to substantiate causal claims, where mechanistic evidence can include a detailed description of what the mechanism is, partial evidence that the postulated mechanism exists, or a plausible mechanism, but at the very least there must be no positive evidence of the absence of a linking mechanism, and both types of evidence are to be considered equally.

While this formulation of the RWT is weak enough to be successful, i.e., every established causal relation should be able to fulfill it; the questions arise once again as to whether it is so weak of a requirement that it is trivial and whether it is really a legitimate version of the RWT.

To answer the first question, in a number of ways the VWRWT protects the integrity of causal claims made in the health sciences. By requiring at the very least that there is no positive evidence of the absence of a mechanism, it forces one to consider what evidence one has of an underlying mechanism. One might argue that these types of considerations are already ingrained in research practice and causal assessment, so much so that having an explicit requirement mandating a search for this type of evidence is superfluous. However, when one reconsiders what the initial RWT requires, it also requires not only that both types of evidence be present, but that they are equally considered. This requirement makes the VWRWT significantly stronger because it carves out a space for the various types of mechanistic evidence to override difference-making evidence. Of course the quality of the evidence in question will constrain this requirement, but it still provides power to mechanistic evidence that has previously been denied (or at least not recognized) by the theoretical construction of evidence-based medicine and other accounts of causality in the context of disease. As such, the VWRWT is not trivial.

Regarding the second question of whether the VWRWT is still a reasonable version of the RWT, the answer is yes. The two main requirements of the RWT are met by the VWRWT: mechanistic and difference-making evidence are both required and the types of mechanistic evidence outlined are capable of indicating confounding and, therefore, can be treated on par
with difference-making evidence. Given the VWRWT’s ability to fulfill both of the original requirements of the RWT, it stands to reason that it should still be considered a version of the original, even if a very distant cousin.

IV. VWRWT & Howick’s Desiderata

After articulating the VWRWT an obvious criticism comes to mind stemming from Jeremy Howick’s two desiderata for sufficiently high quality mechanistic evidence. Howick requires:

1. The knowledge of mechanisms upon which the mechanistic reasoning is based is not incomplete, i.e., there are no obvious gaps in our knowledge of the inferential chain linking the intervention and the patient-relevant outcome.

2. The probabilistic and complex nature of mechanisms is explicitly taken into account when inferring from mechanisms to any claims that a particular intervention has patient-relevant benefit.

The second of Howick’s desiderata is entirely unproblematic and merely addresses the hazards of working with mechanistic evidence in complex systems. The first desiderata, on the other hand, calls into question the legitimacy of the VWRWT’s use of evidence types 2 and 3 because incomplete and plausible mechanisms by definition have gaps in their inferential chains. One can grant, for argument’s sake, that evidence types 2 and 3 do not fulfill Howick’s desiderata and therefore should be dropped from the VWRWT, or at least not valued as highly. For argument’s sake, type 1 can be read as fulfilling Howick’s first desiderata by stipulating that its definition of completeness is wholly inline with Howick’s. However, type 4 evidence presents an interesting case because it does not undertake the project of trying to construct an inferential chain from the intervention to the patient-relevant outcome; it merely indicates the impossibility of such an inferential chain. As such, it appears that this type of evidence falls wholly outside of the domain.

Howick, The Philosophy of Evidence-Based Medicine, 144.
of Howick’s desiderata. Assuming that the bar for positive evidence indicating the absence of a mechanism is high (i.e., it is well-substantiated and trustworthy evidence) it should be taken seriously because it is capable of indicating confounding and, therefore, should be treated on par with complete mechanisms and difference-making evidence.

If it is accepted that Howick’s desiderata would permit the use of mechanistic evidence types 1 and 4, a weak and highly qualified version of the RWT can be defended. Given the concession to Howick that evidence types 2 and 3 are not high quality mechanistic evidence, the final version of the VWRWT would read:

*Both mechanistic and probabilistic evidence are required to substantiate causal claims, where mechanistic evidence may include a complete description of what the mechanism is, but at the very least there must be no positive evidence of the absence of a linking mechanism, and both types of evidence are to be considered equally.*

Stated as such, the VWRWT satisfies both original requirements of the RWT and has been crafted in such a manner that it is both a plausible and defensible account of the evidentiary standards required for causal assessment.

V. Is the VWRWT Useful?

While it may appear that preserving a very weak version of the RWT required an exceptional amount of work for an almost insignificant gain, the reality is that the preservation of the RWT’s two requirements (i.e., (1) the acquisition of both mechanist and difference-making evidence to establish a causal claim and (2) that both types of evidence be treated with equal importance in causal assessment) is largely important to the practice of EBM. Demonstrating the need for one not only to consider both types of evidence in causal assessment but also to consider both types of evidence as of equal importance changes the way one approaches the practice of EBM. If one takes these requirements seriously, one should never evaluate the efficacy of an intervention with
the mindset that RCTs (difference-making evidence) are the gold standard and other types of evidence are not required to substantiate or reinforce the establishment of their causal claims. Because mechanistic evidence is capable of indicating confounding in difference-making evidence, it should, at the least, be considered as a means for verifying the results of RCTs. Additionally, by regarding both types of evidence as equally important in practice, the value of mechanistic evidence is recognized and permits the consideration of mechanistic evidence for other purposes in EBM, such as informing the generation of hypotheses for clinical testing; informing trial design (e.g., what form of placebo we choose will require mechanistic reasoning); informing modifications of trial designs for the same or similar drugs; and providing traditional explanations of biological mechanisms and disease mechanisms (sometimes incomplete).

Additionally, it also encourages one to take seriously the possibility of mechanistic evidence independently establishing patient relevant causal relations between interventions and outcomes (efficacy) due to the fact that the VWRWT outlines those cases in which mechanistic evidence can be considered trustworthy and valuable. Lastly, if one is in the practice of considering mechanistic evidence in all cases, one will also be able to determine when the efficacy of an intervention is supported by both mechanistic and difference-making evidence, in particular, evidence produced by RCTs. This allows for one to have even more reliable method for generating evidence than is currently listed on the evidence hierarchy, something along the lines of “mechanism-supported RCTs.” Taking all of this into account, it appears that the VWRWT has many virtues and should be taken seriously.
VI. Conclusion

This project has analyzed the original articulation of the RWT\textsuperscript{252} and, upon finding its statement imprecise and self-contradictory, interpretations\textsuperscript{253} of its strong and weak versions were also set forth and analyzed. It was argued that the only plausible interpretation of the RWT is Illari’s weak version—the VWRWT. Additionally, after expounding the VWRWT, it was argued that because it successfully met the two requirements set forth by the original RWT it could be considered an honest interpretation of the RWT and, given its minimal concept of mechanistic evidence and its ability to indicate confounding in difference-making evidence, the account also was found not to be trivial. Lastly, the VWRWT was evaluated by Howick’s desiderata for high quality mechanistic evidence and revisions were made to what the VWRWT deemed acceptable mechanistic evidence. In conclusion, a final statement of the VWRWT was presented and put forth as an honest and reasonable interpretation of the RWT. As such, the RWT should be used as a guide for how to engage in causal assessment in the health sciences and as an indication of what types of mechanistic evidence should be taken seriously in evidence-based medicine.


APPENDIX B:
DESCRIPTIONS OF METHODOLOGIES

Randomized Controlled Trials

Randomized controlled trials (RCTs) are often considered the “gold standard” of clinical trials by medical professionals and their methodology (or clinical trial design) is the one most laypersons are familiar with. This is largely due to the success of EBM rhetoric, which has reinforced a categorical understanding of the evidence hierarchy, ranking RCTs as the gold standard at the top. There are many reasons why RCTs are and should be considered the best methodology in some contexts, but there are many caveats to be considered when evaluating their status as a universal gold standard. The primary reason why RCTs have received this status is due largely to the details of their methodology. The benefits of masking and randomization are considered to be the main virtues of RCTs and the reason for their elevated status, though plenty of criticisms can and have been brought against these claims.

The basic definition of a randomized control trial, provided by Sir Michael Rawlins, states,

At its simplest level a randomized controlled trial involves a comparison of the effects of two or more interventions, allocated randomly to groups of contemporaneously treated patients. After a predetermined interval of time the outcomes of the groups are then compared. The interventions are often pharmaceutical products but may include devices, surgical procedures, psychological treatments or any other therapeutic modality.

---

254 Sir Michael Rawlins’ explanations of different trial methodologies are cited a fair amount throughout this chapter because, in comparison to other sources, he explains the methodologies objectively and thoroughly and provides good analysis of them without taking sides in debates regarding the validity and reliability of different methodologies. This means that each methodology is presented in its best light allowing the reader to draw one’s own conclusions. It is acknowledged that Rawlins does have his own biases regarding these issues, but it is the opinion of the author that those biases do not affect the objective approach he takes to explaining these issues. One may argue that presenting the different methodologies in their best light is a demonstration of Rawlins’ biases, however, given that all of the methodologies are employed by respected scientists and social scientists, it is not misguided to acknowledge that there are virtues to each of these accounts that should be presented.

To illuminate the RCT methodology further, random allocation refers to the fact that trial subjects are assigned to an arm (group) of the study through a process similar to flipping a coin, that is, the decision is left to chance and is not decided by anyone with an interest in the study. Some common methods of allocation, such as alternation, have proven insufficient because investigators are able to anticipate which arm of the study the next patient will be assigned to and may, given this knowledge, decide to hold off enrolling a subject for a number of reasons (e.g., they want their patient to receive the experimental treatment or they want to put the healthiest subjects in the treatment arm to achieve the most favorable results). In such cases the allocation is no longer random; the investigators have effectively subverted the randomization process.

The reason randomization is so important to the RCT trial design, and one of the main reasons why it is elevated to such a high status, is the fact that it serves to rule out a number of confounders. Confounders are factors, such as gender, the presence of co-morbidities, or the stage of an illness, which can have a causal effect on the outcome of the study. That is, the existence of such confounders in a study (e.g., an unequal distribution of gender or stage of illness between study arms) can muddy the results of a study. That is, if a study has confounders present, then the study results cannot necessarily be attributed to the intervention being tested because the outcome may have been caused by a confounder. Confounders, in the words of Jeremy Howick, “provide an alternative hypothesis for the result.” Given that the integrity of the study can hinge on the presence of confounders, it is necessary to ensure their absence.

---

256 The choice to use “subjects” rather than “patients” to refer to those individuals enrolled in clinical trials is intentional. Due to the pervasive problem of therapeutic misconceptions (subject confusion between being a subject in a study whose well-being is not the priority of the investigators and being a patient of a doctor whose primary concern is the patient’s well-being) it is the preferred practice of the author not to refer to subjects as patients to keep clear the distinction between receiving medical care and participating in research.

257 Howick, *The Philosophy of Evidence-Based Medicine*, 56.
through good trial design. The most effective tool is randomization, and the most effective methods of randomization are those that cannot be anticipated or intervened upon by investigators. Such methods include making assignments on the basis of whether a subject is born in an even or odd year or by allowing the randomization to be completed by a computer algorithm that cannot be anticipated.

As was explained above, randomization works to rule out confounders. Examples such as gender and stage of illness were given, but these are only examples of known confounders. One may object that we could achieve the same results (i.e., control for all known confounders) by doing away with the randomization process and manually assigning individuals to study arms taking into consideration all known confounders. However, before one objects to such a method on the basis of the types of allocation bias provided above (e.g., investigators making assignments based on scientific or professional biases), there is a more foundational matter to address. Even if investigators were capable of acting as unbiased allocators capable of making perfect allocation decisions based on all known confounders, the problem of unknown confounders would remain. Unknown confounders are those causal factors that cannot be known or anticipated by investigators at the outset of the trial (or possibly ever) and may affect its outcome. An unknown confounder could be something as simple as a dietary interaction with a given medication (e.g., consuming grapefruit can reduce the effectiveness of some pharmaceuticals) or as complex as a physiological interaction with the intervention that only occurs in some subjects. Regardless of what the confounder is, if it is unknown there is no way to

---

258 Some also refer to this kind of bias as “selection bias.” Selection bias is a broader term because it can also refer to a subject’s self-selection to participate in a trial. This is more important when discussing the biases that occur in observational studies. RCTs can be argued to suffer from self-selection bias because it is only those who self-select to participate in research that appear in study populations, however, in observational studies, especially those with historical controls, the self-selection bias refers to those patients who self-select to take a treatment, that is, any treatment at all—a choice of treatment over nothing—which introduces a slightly different kind of bias.
control for it. As such, randomization becomes exceptionally important for balancing unknown confounders because investigators do not have knowledge of the factors and, therefore, cannot control for them in any way other than relying on a random allocation process.\textsuperscript{259}

In addition to randomization, masking (blinding) becomes an important part of the RCT methodology because without it it may be difficult to preserve the benefits of randomization or, at the least, may allow for the introduction of additional confounders after randomization. Masking attempts to remove all biases from the setup and performance of the trial by keeping interested parties in the dark about which subjects are receiving the intervention being tested.

Allocation bias was mentioned in the previous discussion about randomization, however, other forms of bias can interfere with the integrity of a trial by introducing new confounding factors. A primary example of such a bias is performance bias. Performance bias refers to the effects of the expectations or attitudes of patients and dispensers (relevant medical professionals) on the trial outcome.\textsuperscript{260} A common example of performance bias is known as the placebo effect. The placebo effect refers to a subject’s experience of taking a placebo\textsuperscript{261} and simply through the act of receiving an intervention (even though it is inert) experiencing and reporting an improvement in one’s condition.\textsuperscript{262} While the placebo effect is a good example of a performance bias, it is

\textsuperscript{259} This reasoning is largely accepted among medical professionals and EBM proponents. However, some have criticized the effectiveness of randomization at ruling out confounders. In ideal experiments randomization can be repeated indefinitely to improve the likelihood of an equal distribution of confounders, however, it is argued, that it is highly unlikely that all unknown confounders will be balanced through one randomization process in a real world study. Support for this criticism is provided by many instances where the allocation of known confounders is checked between the trial’s arms and are found to be unbalanced. The reasoning stands that if it happens with known confounders there is no reason to believe it wouldn’t happen with unknown confounders.

\textsuperscript{260} For further discussion of performance bias, see: Howick, The Philosophy of Evidence-Based Medicine, 35.

\textsuperscript{261} A placebo generally refers to an inert substance designed to look and taste like the active treatment being provided (e.g., inert pills that (ideally) look and taste identical to the pharmaceutical being tested). Placebos, however, can take many forms. For example, sham surgeries are also considered placebos in trials testing the efficacy of certain surgical procedures.

\textsuperscript{262} The placebo effect is most commonly discussed in the research literature in reference to the use of placebos in clinical trials. However, there are examples of placebo use in routine medical practice. For example, it is reported that some clinicians prescribe antibiotics to patients with viral infections knowing that antibiotics will not aid in curing a viral infection. However, in some cases, the patient feels better simply because they have been prescribed
unlikely that the placebo effect will have a confounding effect on the outcome of the study because it is likely that subjects receiving an active treatment will also experience the positive effects of simply undergoing treatment.

Another example of performance bias that will likely confound the outcome of a trial can occur when a study is not masked and investigators and/or other involved medical professionals monitor more closely the control arm of the study. They might do this because they know that the subjects are receiving a placebo or the current standard of care and they believe those subjects will become sicker without the experimental intervention. This might lead to subjects in the control arm receiving better overall medical care while enrolled in the study in comparison to those subjects in the active treatment arm. As such, the superior medical treatment received by the control arm may influence the study results by indicating that the experimental intervention has a less positive effect than it does.

Masking is the primary tool one has to prevent the introduction of additional confounders through various biases. The most effective masking procedure is referred to as a triple blind. A triple blind refers to the practice of conducting a trial where the investigators/involved medical professionals, the subject, and the individuals analyzing the study data are all kept from knowing which arm(s) of the study are treatment arms and which arm(s) are controls. If a triple blind is meticulously preserved, the likelihood of confounders being introduced through various biases is very low. This fact makes masking just as important, if not more important, than something. How widespread this practice is is unknown. However, such a practice is considered ethically suspect by many.

Studies may also be double blind (investigators/involved medical professionals and subjects do not know which arms are treatment arms) or single blind (subjects are the only parties unaware of which arms are treatment arms). There are of course cases where the active treatment has noticeable or serious side effects such as rashes, vomiting, bleeding, bruising, etc. which make obvious which arm(s) of the study is receiving the active treatment. In these cases it is impossible to preserve the blind in practice.
randomization for preserving the integrity of an RCT’s design. However, in the EBM literature, randomization is commonly hailed as the most important feature of the RCT methodology.

Lastly, an important feature of the RCT methodology is the use of a control. A control refers to an arm(s) of a trial in which the subjects receive either a placebo or the current standard of care to serve as a baseline comparison point for the active treatment arms of the study. Without a control, investigators would be unable to determine whether the changes observed at the end of the study are due to the natural progression of the disease/illness or whether it was the controlled environment of the study setting, etc. If you have a control arm one is able to more reliably attribute the observed effects to the experimental intervention.

Controls are also important because they allow investigators to determine not only that an intervention is effective (i.e., that it causes some positive change in subjects relevant to the outcomes of interest over the course of the study), but that it is as effective or (hopefully) more effective than either no treatment at all or the standard of care. Ideally, all studies should employ standard of care controls because such studies provide the most valuable information for clinicians attempting to make treatment decisions. That is, it is more valuable for clinicians to know whether a new treatment performs better than the standard treatment. Knowing that it works better than nothing may be useful for the aim of advancing scientific knowledge, but it does not help a clinician, ceteris paribus, trying to prescribe the best treatment for one’s patient.

There are a number of different ways in which RCTs can be designed (e.g., a double blind study with four arms consisting of one standard of care control arm and three active treatment arms all with different doses of the experimental intervention; or a triple blind study with two arms: one placebo control arm and one active treatment arm with a dose escalation after

---

265 There are obvious cases in which standard of care controls cannot be used because no standard of care, that is, no treatment, exists for the condition in question.
eight weeks) and the questions they seek to answer can be radically different.\textsuperscript{266} However, the hallmarks of RCTs are the three components mentioned above: masking, randomization, and controls. These components are, as EBM proponents argue, the virtues of RCTs, which make them the gold standard methodology for clinical trials, however most focus on the virtues of randomization.

\textit{Meta-Analyses}

Meta-analyses can be used with methods other than RCTs; however, they can only be performed on those methodologies generating quantitative data. Simply put, a meta-analysis is a pooling or synthesis of quantitative data. Essentially, all data from trials investigating the same research question (e.g., is Wellbutrin an effective aid for quitting smoking?) and using similar methods are pooled and statistical analyses are conducted to determine the effectiveness of the intervention with more precision (i.e., greater statistical power). Such meta-analyses increase the power of the results because the subject pool is larger and, thereby, further reduces the possibility of confounding factors. The smaller a study’s subject pool the less power it has and the more likely it will produce false significant results, that is, misestimate the effect size. Additionally, if there are contradictory results of RCTs and the data is pooled, the meta-analyses done on the pooled data will be more reliable and will produce a more accurate estimate of the effect size.

\textsuperscript{266} There are different “phases” randomized controlled trials can be used for. Most well-known are phases I-III, however some also include phases 0 and IV. Briefly, Phase 0 studies involve background pharmacokinetics and pharmacodynamics; Phase I involves screening for safety (most usually in pharmaceutical trials this involves an investigation of safe dosage); Phase II trials are used to determine the efficacy of the intervention; Phase III involves a final confirmation of the safety and efficacy of the drug; and Phase IV trials are normally conducted after the approval and sale of the drug to further determine side effects, optimal dosage, etc.
A significant problem inherent in the methodology of meta-analyses is that it does not control for bias or poor study design. If a meta-analysis pools the data of 30 studies and 17 of the studies suffer from some sort of bias or they were poorly designed, making their results unreliable, the results of the meta-analysis will be of no better quality than the original studies. A better option, even though the data pool will be smaller, would be to perform the meta-analysis only on the 13 well-designed studies free from bias. As such, it is exceptionally important for those conducting meta-analyses to closely review the methodology of the original studies to ensure they are working with high quality data.267

A further practical problem with this method is that once the meta-analysis has been conducted and its results determined, clinicians consulting the results or those seeking to use the data for a systematic review are unable to determine whether the initial studies used in the meta-analysis were biased unless it is explicitly noted. In practice, this limits one’s ability to determine the reliability of the results produced by meta-analyses.

Lastly, publication bias is a problem capable of seriously affecting the results of meta-analyses. Unfortunately, the practice of many researchers and funding institutions is to curtail studies certain to produce unfavorable results or to delay or not publish the results from completed trials that demonstrate unfavorable results. This practice, known as publication bias, makes it difficult for those conducting meta-analyses to gain access to any negative results. Without the results of unsuccessful trials the results of meta-analyses are more likely to exaggerate positive effects because the data demonstrating negative or null effects is missing.

267 Some disapprove of weeding out studies on the basis of their (poor) methodology because, it is claimed, it introduces unwarranted subjectivity into the process. This objection is viable only if the purported aim of meta-analyses is to analyze all existing data, period. If the aim of meta-analyses is to determine more accurately the effect size of an intervention then it appears that some baseline measure of quality must be invoked and the subjectivity this introduces is wholly justified. Besides, the introduction of subjectivity into methodology is not always bad. For an extended discussion of this issue, see: H. Douglas, Science, Policy, and the Value-Free Ideal (Pittsburgh, PA: University of Pittsburgh Press, 2009). Especially chapters five through seven.
This criticism of meta-analyses is not a direct critique of the methodology because if all trial results (both negative and positive) were published, the results of meta-analyses would actually be reflective of the research that has been conducted, not just that which was published. As was seen in the last critique, one would obviously need the studies to be high quality as well.

Meta-analysis can be a very valuable tool for statisticians and clinicians. However, the quality and completeness of the quantitative data found in the literature will limit the accuracy and usefulness of meta-analyses. Therefore, such analyses should be conducted, presented, and utilized with caution.

Systematic Reviews

Systematic reviews are ranked above RCTs on the evidence hierarchy and are frequently considered one of the best if not the best method for generating high quality evidence.

Systematic reviews provide

a summary of the clinical literature. A systematic review is a critical assessment and evaluation of all research studies that address a particular clinical issue. The researchers use an organized method of locating, assembling, and evaluating a body of literature on a particular topic using a set of specific criteria. A systematic review typically includes a description of the findings of the collection of research studies.  

As one can see from the above definition provided by the Agency for Healthcare Research and Quality (AHRQ), the process used in conducting systematic reviews involves a group of researchers using a predetermined method and set of criteria to assess all the relevant literature pertaining to a particular topic. Missing from the AHRQ definition and included in the definition Sir Michael Rawlins cites is the fact that the “systematic methods [are] selected with a view to

minimizing bias, thus providing more reliable findings from which conclusions can be made and
decisions drawn.”269 This is an important amendment because ignoring the possible ingresson of
bias can lead to unreliable and inaccurate results defeating some of the primary aims of
conducting a systematic review.

By and large, systematic reviews are conducted because of the immense medical
literature and the inability of clinicians (or anyone really) to digest such a large literature within
the constraints of a normal practice—or a normal life for that matter. Systematic reviews offer a
digest of the clinical literature that has already been vetted for quality270 and has gone at least
some way in resolving contradictions among similar studies. Additionally, if a systematic review
includes a meta-analysis of the data, the effect size can be reported with more accuracy. So,
generally, systematic reviews make the clinical research more accessible to those who need to
consult it for guidance in making treatment decisions.

Systematic reviews, when published, include three parts that generally follow the PRISMA
(preferred reporting items for systematic reviews and meta-analyses) guidelines:271

1. Introduction: provides the review question and whether the review is new or an update to
a previous review.
2. Methods: Includes an outline of the main parts of the review, including “eligibility
criteria, search strategies, study selection, summary outcome measures and an evaluation
of the risk of bias.”272
3. Results: quantitative (presentations of the empirical data) and/or qualitative (textual
explanation of the results and their import) analyses of the results are provided.

270 Hopefully. Although not all systematic reviews make exclusions on the basis of quality for the same reasons it is
considered an unacceptable practice in conducting meta-analyses. However, sensitivity analyses can be conducted to
determine if the selections made in the process of conducting the systematic review may have affected the results. If
sensitivity analyses demonstrate that such selections would not have deleterious effects on the outcomes then the
review can be considered reliable in this aspect.
272 Ibid., 47.
By providing the information required in these three sections a clinician should be able to determine fairly quickly (much quicker than consulting the individual studies) the upshot of the literature regarding the particular clinical question being reviewed and what caveats one should be aware of when using the literature. While systematic reviews are practically useful, there are some aspects of the methodology that clinicians should be aware of, including the variability of the results of systematic reviews depending on the selection criteria used as well as the more fine-grained practices of the institution conducting the review.

The Cochrane Collaboration is by far considered the most well-known and well-respected group for generating systematic reviews; however, others do exist and their methodologies differ from group to group. While the methods used by these groups are often made explicit, any variation in the methodology between groups can lead to different results. Additionally, much of the process involves decision-making about what constitutes good research, that is, unbiased and reliable research, and which results are significant. Part of the systematic review process, especially when it includes a meta-analysis, might involve taking insignificant results and judging them significant when considered in conjunction with other insignificant studies.

Unlike meta-analyses, systematic reviews can be used in assessing both qualitative and quantitative results. Attempts have been made to include all empirical evidence generated by any trial methodology in systematic reviews, however this is not standard practice for all systematic reviews. It is without question challenging to combine the results of qualitative studies, observational studies, meta-analyses, etc. in a single systematic review. What often ends up happening is RCTs are the primary source of evidence and other methodologies are referenced in a cursory manner to determine whether the different evidence pools are in agreement or not. However, considering the status of RCTs within EBM, contradictory results from observational
studies, etc., normally do not dissuade clinicians from trusting the recommendations generated by the RCT data.

While there are some cautions particular to the methods of systematic reviews, one must keep in mind that it is largely the quality of the evidence base that will determine the reliability and accuracy of systematic reviews. This is why extreme vigilance in selection criteria attuned to bias is highly important to maintaining the integrity of a systematic review. It should also make one more concerned about the quality of evidence produced in clinical trials and whether the results can and should be trusted.

**Observational Studies**

Observational studies are ranked below RCTs on the evidence hierarchy and are considered the second *best* methodology (after RCTs). The methodology of observational studies differs from that of RCTs primarily in that it does not require randomization for the allocation of interventions. The lack of randomization introduces concerns related to confounding and bias, which RCTs are thought to wholly address—the primary reason they are touted as the *best* methodology. Observational studies, however, are capable of encompassing larger and more diverse patient populations, provide better generalizability—both weaknesses of the RCT—and are much more cost effective than RCTs. Given that each methodology addresses the weaknesses of the other, observational studies and RCTs are commonly considered complimentary methodologies. There are some instances, however, where it is impossible, unnecessary, or even unethical to conduct an RCT and in many of these cases observational studies are an appropriate substitute. Some examples of cases where observational studies would be preferred or necessary in comparison to RCTs include: research on very rare conditions where generating the appropriate statistical power is near impossible due to the small population size; interventions in
which the effect size is very large and, therefore, concerns about confounding and bias do not hold; and interventions where it would be impractical or unethical to conduct an RCT—the best example being surgical research which would require a control arm with sham surgeries.\textsuperscript{273} It is also the case that observational studies are thought to be useful for aiding in generalizing the results of RCTs on similar research questions.

While it is clear that there is a place and a need for observational studies, not all observational studies should be evaluated equally. There are several different types of observational studies and their methodologies differ significantly enough that evidence hierarchies often rank the different types separately and some are excluded completely from the hierarchy. The following will not be an exhaustive summary of these different types, but rather will provide a general outline of their methodologies and differences.

Historical Controlled Trials (HCT)

These studies are not randomized and involve the use of historical controls – that is, patient records/previous case series of patients\textsuperscript{274} in whom the condition and its progression have been

\textsuperscript{273} Sham surgeries are an ethically contentious issue. They are widely condemned; however, some less invasive alternative controls have been employed. For an example see: J.B. Moseley et al., “A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee,” \textit{N Engl J Med} 347 (2002): 81-88. In this study investigators made superficial incisions on the knees of control subjects and simulated the knee surgery received by the subjects in the active arm of the study.

\textsuperscript{274} The use of the term “patient(s)” rather than “subject(s)” is appropriate in describing observational studies because frequently it is the case that the individuals participating in observational studies may not know their records will be used as a control in a future study (i.e., their de-identified medical records will be used) and are receiving standard medical care from a clinician or they are aware that they are participating in a study but their participation is not direct in that they are only authorizing access to their medical records. The choice of whether to try an experimental intervention or opt for standard of care or supportive care is entirely the decision of the patient and can be made in consultation with their doctor who will make decisions based on the best interest of the patient. Because of this the therapeutic misconception is not an issue for observational studies and the term “patient(s)” is in fact a more accurate descriptor of the individual and their involvement.
documented (explicit control) or established knowledge of the natural history of a disease\textsuperscript{275} (implicit control)—for comparison with a group of patients receiving a new intervention. The control and the intervention groups do not receive contemporaneous care and in some cases the historical control arm is constructed by strategically selecting patients who match as closely as possible the profile of the intervention group.

Confounding and bias are serious problems for historical controlled trials because of how the controls are constructed and the inability of researchers to precisely select the control group. Further, for historical controls generated from medical records the necessary descriptions of the condition and/or responses of the patients may not be provided in enough detail if at all. There may also be differences in routine care between the control and intervention groups if there is a significant time lapse between the groups’ treatments. This type of observational study is inferior to other types because of the great risk of confounding and bias.

Concurrent Cohort Studies (CCS)

These studies involve the use of contemporaneous treatment and control groups. These studies differ from historical controlled trials because unlike HCTs, CCSs observe the progress and outcomes of both groups (i.e., treatment and non-treatment) \textit{at the same time}. CCSs can be prospective or retrospective, meaning that the groups can be identified before (prospective) or after (retrospective) the observation period. Retrospective CCSs are not synonymous with HCTs because the control group is not historical, that is, the controls are observed at the very same time as the patients receiving treatment even if the choice to analyze those populations is made after the fact. The contemporaneousness of the group observations make confounding and bias less of

\textsuperscript{275} Syphilis is a good example of an implicit historical control because we have a well-documented natural history of the disease from the Tuskegee Study. It should be noted that this natural history was accomplished in a morally reprehensible way due to the way the Tuskegee study was unethically administered.
an issue than in HCTs because the clinical environment and treatment are much more likely to be comparable. This does not mean that there are no concerns regarding confounding and bias, only that they are greatly reduced.

The methodology of CCSs is most similar to that of RCTs and, consequently, are the specific type of observational study that is considered complimentary to RCTs. The primary difference between RCTs and CCSs is the lack of randomization in CCS methodology. Prospective CCSs especially have very similar trial designs to RCTs: they have a control arm, investigate similar research questions, and have comparable, if not superior, statistical power. Masking is an issue for CCSs, however, it is at least possible to mask their data analyses.

As was mentioned, CCSs still run the risk of confounding and bias due to their lack of randomization, however, there are many cases in which the evidence from CCSs can be considered high quality. According to Rawlins, CCSs provide good evidence when: 1) there is a “biologically plausible basis for the observed benefits and harms”; 2) adjustments are made to the outcomes to account for confounding and bias; 3) the dose-response relationship strongly indicates a causal relationship; 4) a large effect size is observed; and/or 5) similar results have been observed in other studies (especially if the they use different methodological approaches).276

Case-Control Studies (C-CS)

These studies are different from other studies because rather than comparing the effects of an intervention among patient populations (i.e., treated and non-treated), C-CSs compare the

“exposure status to an intervention”\textsuperscript{277} between one group with a certain condition (cases) and another group comprised of healthy normals\textsuperscript{278} (controls). For a higher quality comparison, the normals should be composed of individuals who are “at risk” of developing the condition of interest (e.g., it would be ludicrous to include males as healthy normals in an ovarian cancer or contraceptives study). Further, to ensure the best possible comparison, each case should be strategically matched with a control to rule out confounders. In some cases it is prudent to match a case with more than one control to rule out all of the confounders that can be controlled for. The purpose of such studies is to determine the effects of an intervention.

These studies are generally considered to be inferior to CCSs but are still of high enough quality to be ranked on the evidence hierarchy. There are moderate concerns about confounding and bias with this methodology. If a study fails to match a case with controls well enough, confounding is likely. If a study overmatches cases with controls, there is concern about introducing bias. Additionally, because these studies are concerned with the effects of an intervention, their ability to demonstrate the \textit{benefits} of an intervention is questionable. However, this methodology does exceptionally well in demonstrating the \textit{harms} of an intervention and as such can be very useful.

Case Series (CS) & Case Registries (CR)

Case series extract data from the medical records kept during routine healthcare. Electronic Medical Records (EMRs) are especially useful for this type of study. Case registries involve extracting data from databases created and maintained during research. Because these databases

\textsuperscript{277} Rawlins, \textit{Therapeutics, Evidence and Decision-Making}, 134.

\textsuperscript{278} This term refers to participants in a trial who are healthy and do not have the disease in question.
are typically disease or intervention specific and the records are from medical research, the records and the extractable data they contain are normally more detailed and reliable.

In both CRs and CSs, it is very difficult to control for confounding and bias as the researchers are not involved in the acquisition of data and some of the data may be incomplete or unreliable. The strong benefit to these methods, however, is that they are widely generalizable because they can analyze a large collection of data and the data can be stratified on the basis of age, gender, co-morbidities, etc. These types of observational studies are listed on some hierarchies although they are ranked below all other ranked types of observational studies.

**Case Reports**

This type of study isn’t really a study in the traditional sense. Case reports normally contain a clinician’s write up of an individual patient and a reaction or outcome one had to an intervention. Because a case report is a singular empirical example of a benefit or harm, it is very difficult and potentially dangerous to generalize from a single case report. However, case reports detailing a negative reaction to an intervention can be useful for alerting other clinicians to possible adverse reactions to an intervention. Case reports can and should be conducted for extraordinary interventional outcomes. If enough case reports surface detailing the effectiveness or potential harm of an intervention and a case series develops, this can indicate the need for further investigations regarding the intervention being reported on.

While all observational studies appear on at least some evidence hierarchies in some manner, the most commonly and highly ranked types are concurrent cohort studies and case control studies. The reasoning for this is that historical control trials have a high risk of confounding and bias that makes them unreliable. Case series suffer from similar concerns about
confounding and bias; however their ability to access large amounts of data allows for wide generalizability and mitigation of concerns over confounding and bias. However, because this method primarily relies on medical records constructed outside of a research setting the completeness and reliability of the data still remains questionable. Case reports, while sometimes ranked on the hierarchy, have very few data points as a consequence of their methodology and, therefore, do not have sufficient statistical power to yield generalizable results. However, these methods are able to indicate clinician-observed trends, especially regarding harms, which can indicate to clinicians a need for caution and potentially a need for researchers to test new hypotheses to ferret out whether the harms are real and widely generalizable.

Mechanistic Evidence (Human Models)

The proper definition of mechanisms and mechanistic reasoning is hotly debated in the philosophy of science and it is well beyond the scope of this appendix to enter into that debate. Therefore, it is necessary to employ a general definition of mechanisms that does not commit itself to any particular interpretation of mechanisms. Conveniently, Jeremy Howick provides such a definition: mechanisms “are arrangements of parts/features that…ensure a regular relationship between ‘inputs’ and ‘outputs.’” Mechanistic reasoning, also, involves “an inference from mechanisms” to prediction claims (e.g., daily administration of aspirin reduces one’s chance of having a heart attack).

---

279 Howick, *The Philosophy of Evidence-Based Medicine*, 126.
280 “(allegedly)” has been removed from Howick’s definition. He has included it in his definition because he is skeptical of the ability of mechanisms to indicate stable relations between “inputs” and “outputs.” In certain cases Howick’s skepticism is warranted, however there is no evidence that this skepticism is warranted universally and therefore has been removed from the definition on these grounds.
281 Howick, *The Philosophy of Evidence-Based Medicine*, 128. This particular phrase is borrowed from Howick’s definition of mechanistic reasoning, however the complete definition is too particularized to efficacy claims.
Within EBM, however, there are a host of concerns regarding our understanding and representation of the complex mechanisms functioning within the human body that call into question the validity of mechanistic evidence. As Holly Andersen has astutely observed,

the ordering in the evidence hierarchy reflects the fact that there are enough cases where mechanistic reasoning has failed to improve patient outcomes that the failure is more than merely an occasional exception – it is sufficiently widespread and systematic as to warrant a systematic response.”

That is, there have been enough cases where mechanistic evidence has indicated that a particular intervention would work only to find that the intervention did not work in the way mechanistic reasoning had indicated and consequently harm was done to patients, no benefit was received, or unexpected harmful side effects occurred. As Andersen observes, these errors in mechanistic reasoning have happened frequently enough that it is necessary to “systematically” revamp what we use as evidence in the decision-making process and what status mechanistic evidence deserves in this process.

At this point it is necessary to say a little more about why mechanistic reasoning and mechanisms more generally can be unreliable sources of evidence when predicting the efficacy of an intervention. Bench scientists, such as biologists, engineers and chemists, may wonder why it is that mechanistic reasoning and mechanisms are so poorly regarded in medicine while they enjoy an elevated status in the other sciences.

In the basic sciences, mechanisms are commonly used for two purposes: 1) explanation (i.e., the mechanism can explain the phenomenon in questions) and 2) prediction (i.e., one can predict how mechanisms will respond to an intervention). While the uses of mechanisms are the

---

282 Andersen, “Mechanisms,” 994.
283 An example of such a case would be the old practice of providing oxygen to patients in organ failure. The mechanistic reasoning supporting this intervention was that a cause of organ failure is a decrease in oxygen and therefore supplying oxygen would restore the organs. However, it was later discovered that supplying oxygen to these patients is actually detrimental to the organs. For further information about this case in addition to others like it see: Howick, The Philosophy of Evidence-Based Medicine, 154-157.
same in the medical sciences, what changes is the need to be sensitive to the scope of the mechanism. More specifically, when scientists in the basic sciences construct a mechanism and consult it for explanations and predictions they are focused only on a simple mechanism.\textsuperscript{284} Because the mechanisms are simple they can be examined in isolation without affecting their ability to accurately explain phenomena or predict responses to an intervention. Within the medical sciences, however, it is important not only to understand how a basic mechanism works, but how it will work within the context of a complex mechanism such as the human body. Medical scientists ask applied questions that take as paramount the effect on a patient (i.e., patient outcomes). Because medicine is an applied science, it is necessary for mechanistic explanations and predictions to take into account the effect an intervention will have on the human body as a whole (i.e., a complex system) and not just the isolated effect on an artificially isolated mechanism (i.e., a simple mechanism like “the knee joint”). As such, mechanistic reasoning in medicine runs into the problems of 1) complexity and 2) unpredictability, which make its explanations and predictions less reliable than they are in the basic, unapplied sciences.\textsuperscript{285}

The issue of complexity refers to the fact that a mechanism may lack crucial information regarding how it responds to an intervention when it is situated in a complex causal environment. An example of this would be knee lavage and debridement as a treatment for osteoarthritis.\textsuperscript{286}

\textsuperscript{284} The term “simple” here is not meant to describe the complexity of the mechanism in question, but merely the fact that it is a mechanism extracted from a more complex network of mechanisms or perhaps an independent mechanism that will not need to accommodate or account for the functioning of the other mechanisms it is connected to. For example, a simple mechanism may be a sophisticated lever system built to close a door or engage a break. An example of a complex mechanism would be the circulatory system or the respiratory system. Not only are these highly sophisticated mechanisms, but the functioning of the circulatory system (a highly sophisticated mechanism) can be affected by the (mal)functioning of the respiratory system and vice versa. In these cases, one cannot consider the circulatory completely separately from the respiratory system because a malfunctioning of either can have effects on the other, causing the mechanisms to act in an unpredictable manner.

\textsuperscript{285} Andersen, “Mechanisms.”

\textsuperscript{286} This example is used in Andersen’s discussion of complexity and unpredictability.
After the knee has undergone this intervention and the problem in the mechanism has been solved (i.e., particles have been removed from the knee joint and any tears in the meniscus have been clipped and smoothed) patients still experience pain. That is, even though the apparent problem with the mechanism (i.e., the knee joint) has been fixed, there is still an experience of pain. Why this is remains a puzzle.

Regarding unpredictability, the problem is that even when we have an accurate understanding of different mechanisms, we cannot predict how they will interact with each other under intervention. An example of this is the previous practice of administering prophylactic paracetamol to infants at the same time they are vaccinated against diseases to prevent them from spiking high fevers. The mechanism concerned with bringing fevers down is separate from the mechanism for producing antibodies against disease; however, it turns out that the prophylactic administration of paracetamol retards the body’s development of antibodies, making the infants susceptible to disease. This interaction between mechanisms could not be predicted by understanding the two mechanisms involved. In fact, mechanistic reasoning had suggested that there would be no interaction between the two interventions.

Given that the link between explanation and prediction breaks down when attempting to work with complex biological systems, there is good reason to be skeptical about mechanistic evidence in EBM. However, the absolute lowly status EBM attributes to mechanistic evidence is an overreaction to justified skepticism over its use for determining the efficacy of an intervention. As Holly Andersen states, “there are biological considerations that justify the EBM claim that mechanisms are not a high-quality source of evidence in medicine, with the caveat that

---

287 This example is used in Andersen’s discussion of complexity and unpredictability.
this applies to this particular usage of mechanisms as evidence.” What Andersen is indicating is that there are many high quality uses of mechanisms and mechanistic evidence in EBM. Such uses include informing the generation of hypotheses for clinical testing; informing trial design (e.g., what form of placebo we choose will require mechanistic reasoning); informing modifications of trial designs for the same or similar drugs; externalizing the results of RCTs; and providing traditional explanations of biological mechanisms and disease mechanisms (sometimes incomplete). It should also be mentioned that mechanistic evidence is sufficient for determining the efficacy of some interventions that rely on more basic mechanisms. For example, mechanistic reasoning can accurately indicate the efficacy of the “mother’s kiss” technique for dislodging a foreign object from a child’s nose (i.e., pinch the unobstructed nostril and blow into the child’s mouth).

Lastly, it should be noted that an explanation as to why mechanistic evidence should be regarded as inferior is almost wholly absent from the EBM literature. This indicates that the status of mechanistic evidence has been primarily enforced through the use of rhetoric and has not been subjected to careful and thorough analysis by EBM proponents. There is a growing literature within the philosophy of science that addresses the status of mechanisms in medicine and, in particular, their use in EBM. It is absolutely essential that these exchanges be given more attention by EBM proponents generally.

---

288 Andersen, “Mechanisms,” 996.