Making the Grade

assuring trustworthiness in evidence

Ross Upshur

ABSTRACT  Despite evidence-based medicine’s (EBM’s) significant evolution and maturation from its revolutionary origins to its current form as the preeminent means of practicing medicine, there are still good reasons to be unsatisfied with EBM. This essay explores two important new developments in EBM: recently articulated accounts of the scientific basis of EBM, and the related writings of the GRADE Working Group to create standards for interpretation of the medical literature and evaluation of recommendations. A review of Karanicasos, Kunz, and Guyatt’s (2008) three-step articulation of EBM’s scientific basis demonstrates that the supposed soundness of each principle is not attributable to its scientific status; instead, the normative language of each principle highlights EBM’s grounding in an only partially articulated philosophical framework. The GRADE Working Group’s effort similarly relies on credibility, consensus, and trust in its defense and justification of EBM. These recent developments in EBM reveal that if the clinical research literature is to be informative or foundational to the enterprise of health care, much work needs to be done to secure its trustworthiness and integrity. An agenda for examining trust and trustworthiness in the context of health research is proposed.

The chief danger to philosophy is narrowness in the selection of evidence. This narrowness arises from the idiosyncrasies and timidities of particular authors, of particular social groups, of particular schools of thought, of particular epochs in the history of civilization. The evidence relied upon is arbitrarily biased by the temperaments of individuals, by the provincialities of groups and by the limitations of schemes of thought.

—Alfred North Whitehead (1929)
341. That is to say, the questions that we raise and our doubts depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn.

343. But it isn't that the situation is like this: We just can't investigate everything, and for that reason we are forced to rest content with assumption. If I want the door to turn, the hinges must stay put.

344. My life consists in my being content to accept many things.

— Ludwig Wittgenstein (1969)

While we can acknowledge evolution and growth in the evidence-based approach, I believe that there is still good reason to be unsatisfied with evidence-based medicine (EBM). This article explores new developments in EBM, specifically recently articulated accounts of the scientific basis of EBM and the related writings of the GRADE Working Group. I argue that these developments indicate a new vision of EBM, one radically different from its origins. While EBM continues to evolve, recent writings indicate a shift in orientation, with more focus on aspects of EBM beyond concerns with methodology. In essence, the means by which EBM now seeks ultimate justification is by appealing to its trustworthiness. Yet trust remains an elusive and poorly understood concept in health care. Rather than the generation of more evidence, close attention to trust and trustworthiness of health systems, health-care professionals, and health information is the most pressing concern for the modern enterprise of health care.

The Scientific Basis of EBM

Overt statements describing the scientific basis of EBM have been few. Recently, Karanicolas, Kunz, and Guyatt (2008) stated that:

The scientific basis of evidence-based medicine (EBM) rests on three principles. First, systematic summaries of the highest quality available evidence should inform clinical decisions. Second, wise use of the literature requires a sophisticated hierarchy of evidence. Finally, evidence alone is never sufficient to make clinical decisions; rather, it requires trading off benefits and risks, inconvenience, and costs, and in doing so considering patients’ values and preferences. (p. 1067)

The first principle is a significant carryover from the initial views of EBM. There is an imperative to collect and synthesize all evidence relative to clinical decision making. The chief vehicle for this is the Cochrane systematic review. The historical record indicates significant lags between the discovery of beneficial interventions and their uptake, acceptance, and implementation in practice, and concerns for bias in non-systematic reviews or in single studies. Acceptance of the necessity of systematic summaries of evidence as a justificatory principle requires agreement on what constitutes an acceptable interpretation of evidence in the first place, an as yet undefined concept in EBM.
The second principle introduces two interesting new dimensions: wisdom and sophistication. Inherent in this claim is a belief that interpretive facility, or good judgment, is an important component of EBM, and that it can be reliably identified when it is manifest. Although I agree wholeheartedly with recognizing the importance of wisdom in medical decision making, I fail to see how wisdom is part of the scientific basis of EBM. It is important for the proponents of EBM to look to accounts that have suggested that practical wisdom, or phronesis is relevant to the practice of medicine (Pellegrino and Thomasma 1993; Upshur 1999). Explicit recognition of the interpretive dimensions of engaging the literature, however, entails the introduction of a wide set of skills, many from the humanities and social sciences, that pull EBM away from any straightforward basis in clinical epidemiology. Introducing the notion of the wise use of literature, though entirely appropriate and justified, does not secure the soundness of EBM’s scientific base but actually moves it away from a vision of science rooted in statistical and epidemiological approaches.

Secondly, the introduction of sophistication entails a contrast or differentiation between the true EBM hierarchy and less sophisticated or cruder hierarchies. This introduces concerns similar to those of Brody, Miller, and Bogdan-Lovis (2005) concerning the distinction between crude and sophisticated EBM, and indicates a rupture within the population of those aspiring to claim status as EBM adherents. I will return to this point below in my discussion of the GRADE Working Group, which presumably represents a sophisticated hierarchy. It remains contentious, however, that there is any meaningful sense in which a hierarchy, no matter how sophisticated, is required to secure the scientific basis of EBM. Few sciences conceptualize knowledge in this way. In fact, most well-established sciences maintain a wide variety of standards and techniques for assessing quality of inference and soundness of argumentation.

The idea of a hierarchy itself is not derived from the types of reasoning at the top of the hierarchy (in other words, systematic reviews). That is, when one thinks of the very idea of an evidence hierarchy one must employ conceptual thinking, which is poorly represented in the hierarchy and is at best found near the bottom. A hierarchy of evidence is not something observable in the natural world nor in any sense axiomatic. Ancillary arguments need to be made independent of the hierarchy to explain why evidence can be regarded hierarchically, and why the particular hierarchy proposed by EBM is in any way superior to other candidates. The arguments for the hierarchy rest on concerns for the mitigation of certain biases, such as selection bias and observer bias. It is unclear, however, whether all perspectives engaged in health care are equally concerned with these types of biases. These arguments are not in any straightforward way scientific as understood by EBM but are, more precisely, epistemological.

The third component introduces the now generally accepted claim that evidence itself is not sufficient but requires a set of extra or non-evidential considerations, such as trade-offs between costs, preferences, and values. These consid-
erations operate, presumably, outside the provenance of the hierarchy. In other words, much of the work of making decisions and implementing policies requires the necessary integration of these diverse considerations, most of which are normative. Clinicians wedded to EBM still await instruction on how to achieve this integration.

While these three principles are intended to provide the “sound” scientific basis of EBM, their soundness is not in any way attributable to their scientific status, as there is no straightforward way in which any of these principles are scientific. Each of the principles incorporates evaluative and normative language, making clear that EBM still rests not on a scientific basis, but rather on an as yet partially articulated and largely unexamined philosophical one. Concerns for demonstration of the superiority of outcomes associated with the practice of EBM by virtue of controlled trials or systematic reviews seem to have dropped from consideration of its scientific justification. One would think that this would be the preeminent form of justification, yet it remains an unfulfilled promise (Norman 1999).

**Making the GRADE**

The GRADE (Grading of Recommendations, Assessment, Development and Evaluation) Working Group was convened in 2000 to address concerns regarding the proliferation of hierarchies of evidence and heterogeneity in treatment recommendations. They state on their Web page that: “The working group has developed a common, sensible and transparent approach to grading quality of evidence and strength of recommendations” (GRADE 2008).

GRADE is a striking choice of acronym. It connotes a concern with ranking, evaluating, and recommending certain ways of evaluating the medical literature. It sets out a standard against which all rival hierarchies can be judged as adequate or not. Commonality and sensibility are virtues of its approach, suggesting intersubjectivity, or consensus of views, is integral to the process. One can either make, or fail to make, the grade.

In essence, GRADE seeks to create a uniform, generally agreed upon, explicit, and transparent manner of interpreting the research literature and the recommendations attendant to it. The GRADE group has articulated the rationale for why its approach to evidence appraisal and interpretation is superior to other methods of ranking evidence, offering nine considerations to justify the superiority of its approach (GRADE 2004).

Explicitness and transparency function prominently in the justification of the GRADE approach, and these are important and laudable qualities. Yet they are not self-evident, nor easily derived from principles of EBM. Similarly, concern for judgments, trade-offs between risk and benefits, and incremental costs are not the sorts of things (for the most part) whose justification can come from systematic reviews or randomized control trials (RCTs). Concern for transparency
derivates from more fundamental principles than can be found in statistics. So, once again, when looking for a deeper account for the justification of EBM, one is drawn away from predominantly scientific accounts and towards justifications that draw on ethical considerations. The justification for EBM illustrates its limitations as a scheme of thought. The status of such concepts needs to be argued for and embedded in a justification of GRADE, rather than bluntly asserted.

I have no doubt that GRADE marks an evolution in the scope and concerns of EBM. Important components of the GRADE approach are the recognition of how RCTs can be degraded as evidence; an acceptance that observational studies can provide strong evidence; an enhanced recognition of heterogeneity of perspectives regarding the interpretation of evidence; and a pragmatic emphasis on flexibility in interpreting outcome measures. These are all welcome developments that expand the range of application of EBM considerably and likely reflect an unacknowledged response to previous criticisms.

Yet these advances still leave problematic areas. For example, quality of evidence is now evaluated as either strong or weak: “EBM defines quality of evidence as our confidence in the magnitude of effect estimates for a patient-important outcome. High-quality evidence provides robust results; low quality leaves uncertainty, and the likelihood that best estimates will change with newer high quality evidence” (Karanicolas, Kunz, and Guyatt 2008, p. 1067). On this construal, magnitudes of effect and patient-important outcomes provide the basis of a strong recommendation, and this is supported by claims that further research will not significantly alter this evaluation. Confidence in these results leads to robustness, which is similarly rooted in the “taming” of uncertainty. Low-quality evidence is characterized by high levels of uncertainty and less confidence in its application.

It is questionable whether uncertainty can be hidden behind claims of robustness. The claim here seems to be the untenable one that strong or high-quality evidence, with magnitude of effect stated in terms of some probability statement, is sufficient to secure “certainty.” Yet how the determination of sufficient robustness is determined is unclear. It seems to entail some form of consensus or convention around acceptable bounds of uncertainty, largely without any input or discussion on how these bounds are set. This is a dangerous way to interpret probability claims, as even the strongest evidence does not admit to universal application (Upshur 2005). Furthermore, it obscures and directs attention away from the inherent uncertainties present in even the best evidence and leaves unquestioned the statistical underpinnings of the analysis. I think this shift to regard robustness as a sufficient hedge against uncertainty is a broadly regressive maneuver, as it directs attention away from what Miké (1999) has rightly pointed out as the ineradicable uncertainty at the core of the medical enterprise.

There are further claims in GRADE writings that give rise to concerns: “Quality of evidence is a continuum; any discrete categorisation involves some degree of arbitrariness. Nevertheless, advantages of simplicity, transparency, and
vividness outweigh these limitations” (Guyatt et al. 2008b, p. 925). This is a curious concession. Just how do such considerations outweigh or even balance any arbitrariness? Just how much arbitrariness is tolerable? Indeed, any characterization of such a continuum, discrete or otherwise, entails arbitrariness, or some form of conventionalism. That is, there will be agreement among participants in the process about what is relevant and valued. One reinforced convention in EBM is the continual reliance on p<0.05 alpha levels and 95% confidence intervals. There is no logical justification for this choice, and continued reliance on these measures has significant implications for the adequate interpretation of quantitative data. I have argued elsewhere (Upshur 2001) that these measures create a standard either too restrictive or too permissive depending on the context of the inquiry. Not even R.A. Fisher believed in an invariant standard of statistical significance in each experimental context. As Salsburg (1993) has argued, continued use of such measures lacks logical, mathematical, or statistical justification. Asserting confidence does not answer these fundamental challenges. Furthermore, it is not clear how such considerations as simplicity, transparency, and vividness (again, not necessarily scientific criteria) function to balance arbitrariness. Indeed, it is puzzling to imagine what exactly is entailed by “vividness,” but it does add a dash of panache.

That the weight of evidence may vary according to the context in which it is being interpreted is also explicitly recognized by the GRADE Working Group:

Because governments and public health officials must consider several factors beyond the strength of a recommendation, they may consider that some strong recommendations that are important for individual patients have low priority. These factors—generally of little relevance to recommendations directed at clinicians—include the prevalence of the health problem (higher priority for more prevalent conditions), considerations of equity (higher priority for interventions that tackle health equities by targeting disadvantaged populations), total cost to society (lower priority for interventions with very high total costs), and the potential for improvement in quality of care (higher priority for underused interventions). Thus, if guideline panels are addressing funders or health system managers, they must make transparent the manner in which factors related to prevalence, equity, cost, and improving quality of care influence their recommendations. (Guyatt et al. 2008a, p. 1050)

This paragraph seems to indicate that transparency in terms of considerations of cost, equity, and so forth, are legitimate criteria in decision making outside the clinical domain. Yet this neglects the fact that such concerns are of relevance to individual clinicians and patients, and that micro-level allocation decisions are often taken on such grounds. Furthermore, that factors aside from strength of evidence are determinative in decision making indicates that such factors require a justification that EBM cannot supply. So not only is evidence not enough, but EBM itself is not enough. Claims that the integration of values and preferences
is constitutive to the core of EBM are belied by the lack of significant discussion about how values are understood in EBM—by a failure to recognize and discuss the fact that values may be multiple, conflicting, and in no way as neatly placed in a hierarchy as research evidence.

**The Unexplored Terrain of Trust**

To summarize, it seems that the elements of GRADE and EBM will provide reliable and trustworthy guidance to clinicians, patients, and policymakers in an era of increasingly democratized access to health information, innovations in information technology, a revolution in genomic and molecular approaches to health, and domination of evidence production by large commercial interests. The GRADE group’s appeal to commonality and sensibility posits the existence of groups or organizations who agree on certain core principles and canons of interpretation. This is reinforced by the invocation of the several influential organizations that have endorsed the GRADE approach. In essence, it is a strategy that seeks to establish credibility.

A recent commentary by Montori and Guyatt (2008) warns readers of many false appropriations of the term “evidence-based” and corruptions to the evidence base, and cautions readers that essentially only one EBM brand is the true brand to be trusted. Deviating from adherence to the principles of EBM or embracing illegitimate hierarchies will lead to the medical equivalent of sin: biased decision making.

The most important horizon to be explored beyond evidence is that of trust. Although trust is one of the core values animating health care, it has received scant attention from scholars and practitioners. In this section, I will assert the need to take the ideas of trust and trustworthiness in health care seriously and sketch some potential directions I think should be pursued. Part of this motivation comes from concerns with EBM, and part draws on my own experience as a researcher and clinician. It seems in almost every symposium, workshop, and conference I attend, and in the results of many studies my colleagues and I have conducted, covering diverse topics, trust emerges as an overarching concern. Indeed, at a recent workshop on the future of EBM, trust and trustworthiness figured prominently in discussions. Trust and trustworthiness in health care is a complex issue and will require an account at multiple levels, including such as clinician/patient (micro-level), the intersection of health-care organizations and communities (meso-), and policy making (macro-). Three key areas that need to be addressed are securing the credibility of the evidence base, a philosophical account of trust tailored to health care, and an account of what constitutes trust in medical practice. This is not an exhaustive list, but it should direct thinking to important avenues to explore in order to better understand trust and trustworthiness.
Securing the Credibility of the Evidence Base

If the clinical research literature is to be informative or in some sense foundational to the enterprise of health care, then much work needs to be done to secure its trustworthiness and integrity. As recent research documenting extensive ghostwriting of peer-reviewed clinical trials in the medical literature indicates, there is growing reason to believe that the clinical literature may consist of tainted fruit (Ross et al. 2008). If the RCT is, as Ashcroft (2002) has argued, central to the claims of clinical medicine to be a legitimate, autonomous science in its own right, then aside from concerns about its current epistemic status as argued above, there are important issues related to the actual conduct and publication of RCTs that threaten their integrity. In addition to ghostwriting, the RCT is vulnerable to biases such as outcome reporting bias and publication bias, which render problematic our ability to rely on such studies to inform decisions. If the best estimate of effectiveness is overstated because null studies have not been published, or if the published outcome measure is different from that submitted for review, then no one in the system, including funders, can trust the data or make any confident informed choice whatsoever. As it currently stands, there is no reliable mechanism to secure the medical literature from such influences. Research ethics boards, peer reviewers, journal editors, and readers are not clairvoyant.

Why has it taken defenders of EBM so long to publicly recognize that RCTs are open to manipulation, and that the sobriquet “evidence-based” may be an attractive prefix to multiple interests? This has been apparent for quite some time. One may well ask whether EBM itself is trusted, and whether the practice of EBM exemplifies or manifests qualities and characteristics of trustworthiness.

The answer to this question may be partly negative. For example, in a study colleagues and I conducted examining Canadian family practitioners’ views towards EBM, we found a deep-seated sense of unease with the credibility of research evidence provided or sponsored by pharmaceutical companies (Tracy, Dantas, and Upshur 2003). Concerns pertaining to the credibility and trustworthiness of available evidence are highly troublesome for primary-care physicians attempting to make evidence-based decisions. The results of this study indicate that rank-and-file practicing physicians are skeptical about the quality of evidence derived from pharmaceutical trials. In light of recent events, their skepticism must be regarded as well founded. Similarly a study we conducted to examine Canadians’ views on sharing access to health records indicated mistrust over the idea of government authorities being able to access records, even for projects for which there was agreement that such access was for the common good (Tracy, Dantas, and Upshur 2004). This broad systemic mistrust of EBM must be overcome.
A Philosophical Account of Trust Tailored to Health Care

What is trust and how can trust be justified? Why even think about trust in the context of evidence-based care? It would be helpful if a clear conceptual account of trust existed that would bridge the micro-, meso-, and macro-levels of application.

Philosophical accounts of trust are of some value in helping to clarify matters, but they have not yet been focused on issues in health care. Annette Baier (1986) provides a detailed account of trust, yet her seminal paper ends by concluding that the relationship between two individuals is complex, and it does not inform us how to view collective issues of trust more broadly. A philosophical account of trust would presumably bring greater conceptual clarity and provide rational grounding, but philosophers remain skeptical of the rational justification of trust:

regardless of whether the justification of trust is end- or truth-directed, the exact nature of its justification is puzzling. Should the rationality of trust or distrust be interpreted using an internalist epistemology or an externalist one? Good arguments exist on both sides. And consequently, it is not clear how trust is rationally justified. Neither is it entirely clear what value one could achieve by trusting, given the nature of trust. For example, trust may or may not have moral value depending on whether it can indicate respect for others. (McLeod 2006)

McLeod concludes her account of trust by pointing out how inherently complex, but philosophically interesting and socially important it is to have an account of when trust is warranted. This is a project yet to be completed.

Onora O’Neill (2004) has provided a sustained recent reflection on the determinants of trust in the modern world. O’Neill has moved the account of trust beyond the micro- to the meso- and macro-levels to examine the basis by which organizations and institutions are regarded as trustworthy. She notes that institutions place inordinate emphasis on transparency and openness, which may be inimical to the very notion of trust. That is, claims for the necessity of transparency and explicitness undermine the very function of trust. She writes: “It is very easy to imagine that in a world in which information travels like quicksilver, trust can do the same. It cannot. Placing trust is, I suggest, as demanding today as ever it was in Athens.” Thus philosophical accounts of trust underscore the complexity of the concept. It is clear that further work is required. While EBM has endorsed transparency as an important feature of its current justification, the extent to which this ensures trustworthiness remains unclear.

What Constitutes Trust in Medical Practice

One may also look to the empirical literature and the social sciences for an account of trust. For example, there is a Cochrane systematic review examining the effectiveness of interventions to increase trust in health-care providers (Mc-
Kinstry et al. 2008). The authors observe that patients place less trust in institutions and institutional agents than in individual care providers. In this study, trust is seen as “a global attribute of treatment relationships, one that encompasses subsidiary features such as satisfaction, communication, competency, and privacy, each of which is important in its own right.” They identify trust as a set of behaviors and attributes of physicians, such as thoroughly evaluating problems, understanding a patient's individual experience, compassion, empathy, advocacy, reliability and dependability, communicating clearly and completely, building a partnership, providing appropriate and effective treatment, and being honest and respectful to the patient. They conclude, however, that “Overall there remains insufficient evidence to conclude that any intervention may increase or decrease trust in physicians.”

Thus trust is essential but elusive in nature. Neither philosophical nor empirical accounts adequately ground or explain it. Clearly, as EBM invokes confidence, trustworthiness, and other such credal terms in its justification, a central critical task will be to provide an analysis of trust. I see this as a high priority for any future philosophy of medicine.

While there may be uncertainties about our complete understanding of trust, it is clear EBM’s arguments for its justification are intertwined and constituted by claims for its credibility and trustworthiness. EBM has posed the difficult question of how to cultivate the qualities and characteristics of trustworthiness but has as yet failed to provide the means of inculcating these qualities and characteristics in practitioners. Baier (1986) writes: “Trust . . . is reliance on others’ competence and willingness to look after, rather than harm, things one cares about which are entrusted to their care” (p. 259). It is not surprising, then, that mistrust arises in those contexts where perception of the capacity to do harm, usually in the face of some organizational or impersonal body, outweighs any perception of benefit. So in contexts where relationships, built over time, flourish, so does trust. Trust cannot be secured in systematic reviews or by invoking hierarchies, crude or sophisticated.

Whitehead’s insights regarding narrowness of evidence and the limitations of schemes of thought are worth noting in conclusion. EBM continues to evolve, and one hopes that this will mean an enlargement of its conceptual language and recognition that it need not be so wedded to a narrow conception of science for legitimacy. The language and concerns of GRADE suggest that this is occurring, and one hopes that this expansion of ideas continues. The quotations from Wittgenstein serve as a reminder of how difficult a task it is to question core beliefs. It would be a great consolation to accept things and cease doubting, and this may be the ultimate reason why trust is so important. Yet we cannot be content to accept EBM as it is. EBM may seek to set the GRADE, but it does not, as yet, make the grade.
References

