Topoi
https://doi.org/10.1007/s11245-020-09703-4

What "Evidence" in Evidence-Based Medicine?

Carlo Martini[1,2]

© The Author(s) 2020

Abstract
The concept of evidence has gone unanalysed in much of the current debate between proponents and critics of evidence-based medicine. In this paper I will suggest that part of the controversy rests on an understanding of the word "evidence" that is too broad, and therefore contains the contradictions that allow both camps to defend their position and charge their adversaries. I will argue that reconciling the different meanings of the word 'evidence' in "evidence-based medicine" should help put EBM in its rightful place.

Keywords Evidence-based medicine · Evidence · Expertise

1 Introduction

In 1997, Sackett defined evidence-based medicine (EBM, henceforth) as the "conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients." (1997, p. 3) The definition incorporates the principle that biomedical research about diseases, their etiology, and treatments should be used to aid clinicians' decision-making processes. The concept of best evidence is far from controversial, and Sackett defines it as "patient-centered clinical research into the accuracy and precision of diagnostic tests (including the clinical examination), the power of prognostic markers, and the efficacy and safety of therapeutic, rehabilitative, and preventive regimens" (1997, p. 4).

Since Sackett's paper describing the principles of EBM, there has been much progress in both biomedical research and guidelines translating that research into practical tools for clinicians. But the formulation of guidelines, and the push towards more randomized control trials, systematic reviews and other types of research to support clinical practices have been accused of transforming the practice of medicine into "cookbook medicine", where the guidelines are akin to recipes that clinicians should follow to the letter (see Straus and McAlister 2000).

In the rest of this paper I will follow some of the developments that have taken place in evidence-based medicine since Sackett's paper, and analyze the concept of evidence. I will argue that part of the polarization that we find in the field of medicine and clinical practice, with camps arguing for and against some of the tenets of EBM, rests on an ambiguity about the concept of evidence, which usually remains unanalyzed in the literature. There are at least two meanings of evidence: evidence as sign—e.g., the DNA sample that proves that the killer was related to the victim—and evidence as justification—e.g., the evidence presented in a court of law, which justifies the jury's decision to declare the defendant guilty. The final sections of the paper will elaborate on that distinction.

2 EBM, a Debate Polarized

"Evidence-based medicine" is a relatively new term; it was coined only in the early 1990s but became quickly popular during the late 1990s. The fundamental idea behind EBM is that the practice of medical diagnosis and intervention should be guided by systematic biomedical research. This has not always been the case; for much of its history, medicine was an art, whose principles and techniques were based on experience and passed on mostly through apprenticeship. Goldman and Schafer describe the art of medicine as being

✉ Carlo Martini
uni.c.martini@gmail.com

[1] Faculty of Philosophy, Vita-Salute San Raffaele University, Milan, Italy

[2] Centre for Philosophy of Social Science, Department of Political and Economic Studies, University of Helsinki, Helsinki, Finland

Published online: 29 March 2020

"guided by millennia of common sense" (2012, p. 1). Personal knowledge and experience have a primary role in the development of science and technology: Knowledge cannot always be codified and passed on as explicit information, and direct expert-to-expert knowledge transfer is an integral part of the progress of science (see Polanyi 1958, Collins and Evans 2008).

Claridge and Fabian (2005) describe an early period of evidence-based medicine as dominated by historical and anecdotal accounts, followed by a "renaissance" period in which practitioners started to record their observations in journals, and textbooks became more prominent, often in the form of treatises. Despite advances in the effort of codifying medical knowledge, personal clinical experience remained a fundamental source of knowledge in medical practice for much of the twentieth century. In the second and late part of the century, research on cognitive biases and heuristics started to question blind reliance on expertise, experience, and personal knowledge for decision making (see Kahneman et al. 1982). Already in the 1950s, Paul Meehl questioned the accuracy of expert judgment when it comes to the prognosis and treatment of psychological and psychiatric disorders (see 1954), and subsequent research highlighted the shortcomings of expert clinical judgment when taking decisions based on complex and uncertain information (see Blumenthal-Barby and Krieger 2015). The developments in the Biases and Heuristics program since the work of Meehl led researchers to look for better ways of systematizing and categorizing medical knowledge, and that, in turn, led to the recent developments of the evidence-based medicine movement: Sackett's definition of EBM, mentioned in the introduction, the work of the Cochrane Collaboration, and a number of other initiatives aimed at getting the principles of EBM in touch with clinical practice (see Claridge and Fabian 2005).

Despite the good intentions of its first proponents, EBM is viewed as a near cult by a significant number of theoreticians and practitioners in contemporary medicine. A cult is a system of beliefs with little or anyway partial rational grounds, a system held together principally by habit, faith, or other sociological factors, and a cult has acolytes as well as detractors. Acolytes might be in the cult because of personal interests, financial interests, but also by pure indoctrination and irrational faith to the tenets of the cults. Detractors might have personal grievances with the cult, as well as good reasons and concerns with both the tenets of the cult and the behavior of the acolytes.

This picture is most likely an exaggeration of reality, but it is hard to deny that many scholars and practitioners tend to highlight EBM's shortcomings in ways that make it look like a cult. The number of charges against EBM include: (1) EBM provides, whether intentionally or not, a cookbook approach to medicine (see Feinstein and Horwitz 1997); (2) EBM focuses almost exclusively on randomized controlled trials (see Cartwright 2018); (3) EBM provides too easy a framework for policy-makers and health care providers to constrain clinical practices and even refuse treatments that are not included in the approved guidelines (see Feinstein and Horwitz 1997), which are heavily biased towards randomized control trials (RCTs); and finally, among the peculiarly philosophical challenges, (4) EBM makes strong metaphysical assumptions and relies on a flawed positivist methodology (see Anjum 2018).

While the metaphor of the cult cannot be taken too seriously—both EBM proponents and its detractors have serious theoretical, methodological, practical, and even metaphysical concerns—the polarization between the two camps should not be minimized; it has been highlighted for over a decade (see Timmermans and Mauck 2005) and the rhetoric has not been toned down ever since. This is bad for medicine, and bad for science: Probably, the last thing a patient would want to hear are doctors and nurses bickering about which practice is most supported by evidence.

Finding the roots of the controversy is a step towards resolving the differences, and in the next sections I will suggest that part of the controversy rests on an understanding of the word "evidence" that in the debate between proponents and detractors of evidence based medicine is both unanalyzed and too broad, and therefore contains the contradictions that allow both camps to defend their position and charge their adversaries. As mentioned at the beginning, however, the evidence-based movement in medicine was born with good intentions. Sackett (1997) presents the concept of EBM as the attempt to integrate systematic research and expert judgment, starting from the observation that expert judgment is flawed (see Faust 1984) and needs help from systematic research and data. I will argue that reconciling the different meanings of the word 'evidence' in evidence-based medicine should help put EBM in its rightful place.

3 The Perils of Judgment

In his 1997 article Sackett poses a distinction between clinical expertise, which directs the clinician's actions on individual patients, and external evidence, which should inform those actions. Without external evidence, Sackett claims, "practice risks becoming rapidly out of date, to the detriment of patients." (1997, p. 3). Tracing the evolution of EBM, Swanson et al. (2010), highlight the problem of misinformed practice as one of the reasons why we need external evidence in medicine. They claim that "unproven practices are still being recommended without evidence that they actually improve outcomes" (Swanson et al. 2010, p. 287). Several of the studies on biases and errors in clinical judgment carried

out in the wake of Meehl's work were quite damning. They exposed the limitations of expert judgment when judging a clinical case and often highlighted the superiority of actuarial/statistical procedures for diagnostic purposes. Faust (1984) cites several studies showing that human judgment in clinical settings can be dangerously close to a lucky guess, with some of the studies reporting success rates in diagnosis close to a random assignment (see Faust 1984, pp. 40–43).

Experiments aimed at testing the judgment of a doctor in idealized settings are not directly representative of the complexity of real clinical cases. Therefore, even if doctors' judgment abilities fail in hypothetical clinical reasoning tasks, that might not be an indication that they would fail in the real world of patients. The problem is a well-known and contentious issue in the Heuristics and Biases literature: Do the biases observed in experiments exist in the natural environment? Or are they a by-product induced by the testing scenario, the framing of the questions, and so on? (see Gigerenzer et al. 2011) Blumenthal-Barby and Krieger (2015) provide a systematic literature review on cognitive biases in medical decision-making. Their conclusions are more cautious, compared to the literature reviewed in Faust (1984), and their own review of the decade of studies between the 1980s and the 2010s contains 213 studies, of which 49 are studies about real-world medical decisions, rather than hypothetical decisions made based on vignettes or idealized scenarios (Blumenthal-Barby and Krieger 2015, p. 541). Real-world studies should be taken seriously, and even though the authors suggest that additional future research is needed to study actual medical decision-making, their review is a clear indication that the phenomenon of biases in medical decision-making is real, and not just a byproduct of research on biases.

There are at least two issues with relying on expert judgment in medical practice. Firstly, not only common sense but also psychological research tells us that judgment in humans is sticky: Once an opinion or practice becomes consolidated, we do not change our mind easily, especially if the cognitive load of choosing the alternative is high. The phenomenon comprises a number of biases, among which the status-quo bias (see Fernandez and Rodrik 1991) and the confirmation bias (see Klayman 1995) and, while the debate is ongoing as to how to measure cognitive biases, the general tendency is clear: human judgment does not change easily in the light of new evidence. The second problem is that humans do not navigate easily in the complex world of probabilities and uncertainties: it is also known from experiments in cognitive psychology that humans make even very basic mistakes when considering probabilities, from ignoring base rates (Bar-Hillel 1983), to confiding too greatly in their own predictive powers, a phenomenon called overconfidence (Croskerry and Norman 2008). The two problems combined make the possibility of fallacious clinical reasoning even greater. Hamm (1996) reports on real cases in which doctors fail to take base rates into account (base rate fallacy), or whose judgment is affected by the order of the presentation of the evidence (order effect: primacy/recency). According to Hamm, doctors should know probabilities in order to manage the patient, and he concludes that "it is wrong to deny that physicians have a problem of base rate neglect here, as well as a more general problem of inaccurate probabilistic inference." (1996, p. 26).

The problems are not resolved by presupposing that expert judgment—i.e. the judgment of trained professional clinicians—must be better than the judgment of the average reasoner. David Faust and others have amply corroborated the observation that error in judgment affects experts as well as laypeople (see Faust 1984). Biomedical research moves too fast a pace to let us believe that healthcare professionals can keep up with the amount of evidence being churned out of professional journals in the almost endless selection of medical subfields. Here, too, the numbers are clear, the pace of research is so fast that in the last decade researchers have turned to data-mining algorithms to collect information from published scientific articles (Yoo et al. 2012). Nowadays, even more so than when Sackett was writing his piece, it seems obvious that unaided medical judgment is insufficient for proper patient care. Expert judgment is fallacious, and initiatives to collect and systematize evidence in order to formulate guidelines that can help the clinician must be welcomed for patients' sake.

4 Combining Expert Judgment and Statistics

One of the hard problems for EBM is the fact that we need to translate knowledge from the population-level of most EBM studies to the level of individual patient-centered care. Anjum et al. argue that drawing inferences from population level to individual level is a fallacious strategy because the "most relevant sub-group is always the N-of-1 group." (2015, p. 11) Most critics of EBM are quick to point out that what is needed is more judgment, more clinical expertise and more contextual and patient-centered reasoning. "Judgment and interpretation as part of the practice have been set to one side in favor of a highly empiricist account" (Kelly 2018, p. 1164). Anjum cites "trust in clinical judgment" as one of the components necessary to refocus the wrong turn in EBM (2018, p. 1131). In Cartwright's framework, using a diversity of methods and sources of evidence for establishing causal claims still needs judgment for the amalgamation of heterogeneous, and possibly discordant, evidence (see Cartwright 2018). The critique of EBM can take at time a romanticized turn: Feinstein and Horwitz claim that "advocates of EBM may often be diverted from the bedside to the

library or computer terminal" (1997, p. 533). This depiction of medicine, while appealing in its sentimentalist tone, does not capture the fact that both the library and the terminal are necessary tools for dealing with the complexity of factors involved in medical diagnosis and treatment. In short, appeal to judgment needs to be taken with a grain of salt. While I have argued, elsewhere that judgment is necessary in decision making (see Martini 2014), what we need is not judgment simpliciter, we need structured judgment from selected sources.

The call for more judgment and less strict guidelines is not peculiar to critics of EBM. In Swanson et al. (2010), an article supporting the principles of EBM, the role of expertise is stated explicitly: "clinical expertise and patient values and preferences are key elements of EBM and are equally important in clinical decision making." (2010, p. 291). Burns et al. (2011), a widely cited article that stresses the importance of scales of evidence, rejects the role of RCTs as the golden standard of EBM and highlight the need for cautions interpretation and understanding of all levels of evidence. In short, it seems that both EBM proponents and critics share the principle that scales of evidence should be an aid to, and not a replacement of clinical expert judgment, the latter retaining its central role in medical practice.

But the question remains: in what way can scales of evidence be of help to expert judgment? If the pitfalls of relying on judgment were not enough, a further problem exacerbates the situation. Let us imagine a scenario in which guidelines served as mere guidelines, indeed a piece of advice, to be used in the appropriate context, rather than a rule in a cookbook. The clinician could follow the advice of the guideline if the context is appropriate, and deviate from the guidelines when they think they should do so. This strategy is called "selective defection strategy" and according to Bishop and Trout (2005, p. 53) they tend to underperform the strict use of rules. In contexts in which guidelines have been shown to overperform human judgment, according to Bishop and Trout defecting from the rule (a statistical prediction rule) should be done only under very exceptional circumstances. The problem is that humans tend to defect from a rule more often than they should: "Typically, subjects find more broken leg examples than there really are." (Bishop and Trout 2005, p. 47).

The broken-leg problem is often cited as a limitation of statistical reasoning. Gawande's popular science book summarizes it well: "A statistical formula may be highly successful in predicting whether or not a person will go to a movie in the next week. But someone who knows that this person is laid up with a broken leg will beat the formula. No formula can take into account the infinite range of such exceptional events. That's why doctors are convinced that they'd better stick with their well-honed instincts when they're making a diagnosis." (Gawande 2010). It is telling that Meehl himself, a staunch advocate of statistics over judgment, was aware of broken-leg problems. Meehl (1957) explains that once we identify the possibility that the case at hand could be an exceptional one, the question turns to whether that case belongs to the class of exceptions or to the class of cases that are captured by the statistics. In other words, if we do not know whether the person has actually broken her leg or not, how likely is it that she has? According to both critics and proponents of EBM, clinicians should follow guidelines, except on special occasions, i.e., when there is a broken leg. But at this point the issue has shifted to the question of whether identifying an exceptional case is something that clinical judgment is good at. Whether the case is truly exceptional, and not just one of the cases captured by the statistics, is itself a judgment, and human reasoners tend to fail at this task too (see Bishop and Trout 2005).

One must note here that it is different to *know* that we are in a broken-leg case, and to *have a hunch* that the case we are dealing with is an exceptional one, which cannot be captured by the statistics. The former instance is an easy one, and clinicians are better suited at addressing those cases than statistical rules. The latter situation is more complicated: how often do broken leg occur for a given class of problems, say, a medical decision scenario? In sum, reconciling guidelines and judgment is further complicated by the fact that human judgment does not mix well with rule-based judgment, and even the strategy of rule-defection under appropriate circumstances runs foul of the same biases of judgment described in the previous section.

5 Evidence Of and Evidence For

EBM is usually described as the "conscientious, explicit and judicious use of current best evidence […] in making decisions about the care of individual patients." (Swanson et al. 2010, p. 286) The concept of evidence is central to epistemology and usually refers to "that which justifies beliefs". Bertrand Russell thought that evidence is "sense data", that is, belief-justifying mental items, and in a number of classic Gettier-style examples, evidence is what allows a human subject to believe something rationally. For instance, the bloodied footprint on the scene of the murder is what justifies Sherlock Holmes in believing that the murderer was wearing boots, and not sandals; and traces of DNA on the victim's body might help the jury reach a verdict in a murder trial. According to Bayesians, additional evidence helps us strengthen or weaken our beliefs, even when we form them in conditions of limited evidence. In this sense, evidence is evidence for someone, of something. The passage from an item of evidence to a belief is mediated by a number of factors, like capacity of reasoning, context, and prior beliefs.

The concept of evidence as what justifies beliefs is not the only one that we find in the philosophical literature. For many philosophers and methodologists of science, evidence is the courtroom of science: it is public in nature, it is in front of everyone's eyes, and it is in virtue of it that we can settle scientific questions transparently. Proponents of the latter concept of evidence include Rudolf Carnap, who saw the goal of philosophy as developing a "theory of evidence that will enable scientists to settle disputes […] over whether, or to what extent, putative evidence confirms a hypothesis" (Achinstein 2010, p. 36). Most contemporary accounts of evidence take evidence to be evidence for something, without accounting for the human factor. Such are the accounts by Achinstein (2010) and Woodward (2003), and the probabilistic concept of evidence (see Bovens and Hartmann 2004).

Achinstein writes "[…] let me say that the notion of evidence I am concerned with is an objective, not a subjective, one: whether *e* is evidence that *h*, and how strong that evidence is, does not depend on what anyone believes about *e*, *h*, or their relationship" (2010, p. 36). Similarly, albeit implicitly, Woodward endorses a notion of evidence that does away with the subjects who process such and such evidence and come to such and such conclusions: "The existence of a correlation between X and Y that persists under the interventions specified in the antecedent of this counterfactual is in turn evidence that the counterfactual is true" (2003, p. 105).

The Bayesian account of confirmation, a popular one in current philosophy of science, affirms that if *E* is evidence for a hypothesis *H*, then it is more likely that *H* is true, given *E*[1] (see Bovens and Hartmann 2003). Evidence is thus related to truth, and in science in particular, evidence is an indicator of causal relations. For instance, drug A will be effective at treating condition C, because of a causal relation between the chemicals presents in the drug and the response of the organism to those chemicals.[2] Evidence—i.e., a marker—of this causal relation could be, under appropriate conditions, the phenomenon itself: when we take the drug, the symptoms of the condition disappear. The analysis of causes that starts from what we observe to what we can infer is called causal analysis (see also Cartwright 2018). In the latter sense, evidence is evidence of something. As Hacking famously put it, according to the scientific and probabilistic understanding we can talk of "evidence of things" (see Hacking 2006, chapter 4) and evidence is a sign of something (e.g. a causal relation).

6 Evidence in Evidence-Based Medicine

Part of the disagreement between supporters and critics of EBM rests on the failure to take notice of the differences in uses of the word "evidence". Medicine is both a practice, and a science; as a science, medicine is concerned with causal relations between conditions (e.g., illnesses), and a range of factors like bacteria, viruses, chemicals, nutritional elements, and so on. Most science works by categorizing phenomena; that is, by finding common features between classes of phenomena. We may say that each token-bacterium belonging to the salmonella genus is in an ontological sense unique, but we classify them as belonging to the same genus. Moreover, we find that the salmonella bacteria tend to cause an infection called salmonellosis, which displays a certain range of symptoms and responds to a certain class of antibiotics.

According to a certain interpretation of Aristotelian thought, knowledge can only be knowledge of universals (see Leszl 1972). In this sense, medicine as a science is knowledge of diseases, understood as general categories. It is most likely that a *disease A* will differ slightly from *individual X* to *individual Y*, but for the purposes of science it is useful to categorize diseases because, under the appropriate category, both *individual X* and *individual Y* will respond to the same treatment. This is not a trivial point to make: While it is obvious that patients are particulars, it is useful to categorize diseases as universals, in the same way that we categorize almost identical chemical compounds as the same drug, because same disease responds to the same drug in idealized conditions. It is obvious that the principle does not hold in general: in an organism it is not generally true that the same disease responds to the same drug. But for the purposes of scientific research and categorization the principle is useful. In this scenario, the evidence we are looking for is evidence of a cause: the drug interacts causally with the disease, and produces a desired effect. In this sense we are dealing not with individuals, but rather with populations, we are interested in experimental control, average treatment effects, and so on.

The study of these phenomena, as types, is the typical business of biomedicine, which categorizes types—of organisms, diseases, chemical compounds, etc.—and studies interactions among types by means of a range of methods, each of which has advantages and disadvantages. The modern tendency is to put such advantages and disadvantages on a scale, i.e., a scale of evidence. Scales of evidence have been amply criticized, since even systematic literature reviews—the type of evidence that usually lays at the top

[1] More precisely, according to Bayesian Confirmation Theory, E is evidence for a hypothesis H, *if and only if* P(H|E) > P (H) and E is evidence against hypothesis H, if and only if P (H) > P(H|E).

[2] This is clearly not the only possibility: The drug might be effective because of a placebo effect, but for simplicity of the example we can bracket this and other possibilities.

of a scale—can be biased if there are file-drawer problems, a notable source of bias in the statistics, which can cause the effect of an intervention to be unjustifiably magnified (see Rothstein et al. 2005). All sources of evidence, from randomized controlled trials, to case reports and, as argued above, expert opinion, can be biased. But clearly there will be classes of problems for which one bias is more or less likely. In cases where the decision we have to take requires complex probabilistic information, we know that human reasoning is not well tuned to handling incomplete and probabilistic data—for instance, humans subjects consistently fail at accounting for base rates when calculating probabilities (Pennycook and Thompson 2016). It seems obvious that in those cases we would like doctors to receive help in the form of guidelines when base rates are involved, rather than letting their judgment rely on guesses and hunches.

As a practice, medicine is concerned with diagnoses and treatments of particular cases, not of universals. In Aristotelian philosophy, this is a different realm of knowledge, it has to do with practice rather than theory (see Jonsen and Toulmin 1988). The realm of practice requires the application of general principles to specific cases: "practitioners appeal to universal atemporal theories chiefly for the help they may give in dealing with practical problems arising here and now." (Jonsen and Toulmin 1988, p. 32) The focus of practice is different than the focus of science, and so are the concepts of evidence that we can apply to the two cases. As a practice, doctors need evidence in order to diagnose and treat a patient. Diagnosis and treatment correspond to the phases of believing and acting: we can believe and act in a rational way if we have evidence of the right kind. The focus of both diagnosis and treatment ought to be the individual patient, even when we treat a population, we are not necessarily interested in the collective, but in the individuals. Accordingly, getting evidence for treatment and diagnostics will involve getting "knowledge about the individual patient and which causally relevant factor might affect the interaction with the treatment." (Anjum 2018, p. 1129).

Practice requires not only the application of theory to specific cases, but also the need to negotiate between opposing theories, the weighing of (possibly discordant) evidence, and the consideration of values that may be attached to the specific problems. Evidence, in practical considerations, is not evidence of something, simpliciter; but rather, evidence for someone, of something, and, as such, it requires an argument—that is, "a network of considerations, presented as to resolve a practical quandary." (Jonsen and Toulmin 1988, p. 32) Arguments, in turn, require judgment (only very few argument can be explicated completely and turned into formal ones), and judgment requires expertise.

If we understand properly the relation between the concept of evidence in argumentation theory—evidence is evidence for someone, of something—and the concept evidence as a sign of something—for instance, a cause—then some of the disagreement between proponents and detractors of EBM could be put to rest, or at the very least mitigated.

7 Conclusion

In this paper I have started from the observation that evidence-based medicine is a polarized field (Timmermans and Mauck 2005). On the one hand, the autonomy of clinical judgment is fundamental in the practice of medicine, where the focus is on individual patients. On the other hand, autonomy cannot translate into a pretense for omniscience, and medicine, as a science, needs reliable knowledge for guiding clinical practice. Human judgers are fallible, and so are doctors. In order to provide the best care to patients, physicians have an array of tools to improve their decision-making: the several methodologies in use in biomedicine, joined with the knowledge of the limitations of each of those methods. As I have illustrated in Sect. 4, however, it is not easy to combine general knowledge, in the form of statistics and guidelines, with the judgment of the clinician focused on individual patients. These difficulties, and the possible polarization of antithetic fields, has led to extremes that are neither useful for practice nor fruitful for theory. On the one hand, the idea that clinical judgment should have the exclusive epistemic authority over diagnosis and treatment of a patient. On the other hand, the idea that doctors should follow clinical guidelines blindly and a-critically.

In this paper I have tried to illustrate how a more profound comprehension of the different concepts of evidence already present in much philosophical literature can help better understand the root of the disagreement and possibly help unify the field. At least some of the qualms with evidence-based medicine—the idea that cook-book medicine could replace the judgment of experts and practitioners—could be put to rest, if we start from the observation that there are at least two senses of evidence that ought to be considered: evidence of something, in medical sciences, and evidence for someone of something, in medical practice.

Acknowledgements Open access funding provided by University of Helsinki including Helsinki University Central Hospital.

Funding This study was funded by the Finnish Cultural Foundation (Grant Number 0116947-3).

Compliance with Ethical Standards

Conflict of interest Carlo Martini declares that he has no conflict of interest.

Research Involving Human Participants and/or Animals This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Achinstein P (2010) Evidence, explanation, and realism. Essays in philosophy of science. Oxford University Press, Oxford

Anjum RL (2018) What is the guidelines challenge? The CauseHealth perspective. J Eval Clin Pract 24(5):1127–1131

Anjum RL, Kerry R, Mumford SD (2015) Evidence based on what? J Eval Clin Pract 21:E11–E12

Bar-Hillel M (1983) The base rate fallacy controversy. Adv Psychol 16:39–61

Bishop MA, Trout JD (2005) Epistemology and the psychology of human judgment. Oxford University Press, Oxford

Blumenthal-Barby JS, Krieger H (2015) Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. Med Decis Making 35(4):539–557

Bovens L, Hartmann S (2003) Bayesian epistemology. Oxford University Press, Oxford

Burns PB, Rohrich RJ, Chung KC (2011) The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg 128(1):305–310

Cartwright N (2018) What evidence should guidelines take note of? J Eval Clin Pract 24(5):1139–1144

Claridge JA, Fabian TC (2005) History and development of evidence-based medicine. World J Surg 29(5):547–553

Collins H, Evans R (2008) Rethinking expertise. University of Chicago Press, Chicago

Croskerry P, Norman G (2008) Overconfidence in clinical decision making. Am J Med 121(5):S24–S29

Faust D (1984) The limits of scientific reasoning. University of Minnesota Press, Minneapolis

Feinstein AR, Horwitz RI (1997) Problems in the "evidence" of "evidence-based medicine". Am J Med 103(6):529–535

Fernandez R, Rodrik D (1991) Resistance to reform: status quo bias in the presence of individual-specific uncertainty.Am Econ Rev 81(5):1146–1155

Gawande A (2010) Complications: a surgeon's notes on an imperfect science. Profile Books, London

Gigerenzer GE, Hertwig RE, Pachur TE (2011) Heuristics: the foundations of adaptive behavior. Oxford University Press, Oxford

Goldman L, Shafer AI (eds) (2012) Approach to medicine, the patient, and the medical profession: medicine as a learned and humane profession. In: Goldman's Cecil Medicine, 24th edn. Elsevier, London, pp 2–3

Hacking I (2006) The emergence of probability: a philosophical study of early ideas about probability, induction and statistical inference. Cambridge University Press, Cambridge

Hamm RM (1996) Physicians neglect base rates, and it matters. Behav Brain Sci 19(1):25–26

Jonsen AR, Toulmin S (1988) The abuse of casuistry: a history of moral reasoning. University of California Press, Berkeley

Kahneman D, Slovic P, Tversky A (eds) (1982) Judgment under uncertainty: heuristics and biases. Cambridge University Press, Cambridge

Kelly MP (2018) The need for a rationalist turn in evidence-based medicine. J Eval Clin Pract 24(5):1158–1165

Klayman J (1995) Varieties of confirmation bias. Psychol Learn Motiv 32:385–418

Leszl W (1972) Knowledge of the universal and knowledge of the particular in Aristotle. Rev Metaphys 26(2):278–313

Martini C (2014) The role of experts in the methodology of economics. J Econ Methodol 21(1):77–91

Meehl PE (1954) Clinical versus statistical prediction: a theoretical analysis and a review of the evidence. University of Minnesota Press, Minneapolis

Meehl Paul E (1957) When shall we use our heads instead of the formula? J Couns Psychol 4(4):268

Pennycook G, Thompson VA (2016) Base-rate neglect. In: Pohl RF (ed) Cognitive illusions. Intriguing phenomena in judgement, thinking and memory. Psychology Press, Hove

Polanyi M 2012[1958] *Personal knowledge.* Routledge, London

Rothstein HR, Sutton AJ, Borenstein M (eds) (2005) Publication bias in meta-analysis. In: Publication bias in meta-analysis: prevention, assessment and adjustments. Wiley, New York, pp 1–7

Sackett DL (1997) Evidence-based medicine. Semin Perinatol 21(1):3–5

Straus SE, McAlister FA (2000) Evidence-based medicine: a commentary on common criticisms. CMAJ 163(7):837–841

Swanson JA, Schmitz D, Chung KC (2010) How to practice evidence-based medicine. Plast Reconstr Surg 126(1):286–294

Timmermans S, Mauck A (2005) The promises and pitfalls of evidence-based medicine. Health Aff 24(1):18–28

Woodward James (2003) Making things happen: a theory of causal explanation. Oxford University Press, Oxford

Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, Hua L (2012) Data mining in healthcare and biomedicine: a survey of the literature. J Med Syst 36(4):2431–2448

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.