Epidemiological Evidence: Use at Your 'Own Risk'?

Jonathan Fuller*†

What meaning does epidemiological evidence have for the individual? In evidence-based medicine, epidemiological evidence measures the patient's risk of the outcome or the change in risk due to an intervention. The patient's risk is commonly understood as an individual probability. The problem of understanding epidemiological evidence and risk thus becomes the challenge of interpreting individual patient probabilities. I argue that the patient's risk is interpreted ontically, as a propensity. After exploring formidable problems with this interpretation in the medical context, I propose an epistemic reinterpretation of individual patient probabilities as credences. On this view, epidemiological evidence informs medical uncertainty.

1. Problem of the Meaning of Population Evidence for the Individual. Patients are not statistics. This statement is trite but true and often underappreciated in an era in which medical care is driven by evidence from epidemiological studies. The puzzle of what bearing group data have on individual patients is not new. It was hotly debated by physicians, physiologists, and statisticians in the nineteenth century at the dawn of medical statistics (Matthews 1995). In the twentieth century, clinical trials pioneer Austin Bradford Hill reflected that clinical trial results present "a *group* reaction. They show that one group fared better than another, that given a certain treatment patients, *on the average*, get better more frequently or more rapidly, or both. We cannot necessarily, perhaps very rarely, pass from that to stating exactly what effect the treatment will have on a particular patient" (1952, 117).

*To contact author, please write to: Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA; e-mail: JPF53@pitt.edu.

†Thanks to Alex Broadbent, Steven Goodman, Stephanie Harvard, Stephen John, Richard Kravitz, David Spiegelhalter, Jacob Stegenga, and Mark Tonelli, as well as audiences at the 2018 Philosophy of Science Association biennial meeting in Seattle, the University of Toronto, the University of Bordeaux, and the University of Pittsburgh for helpful feedback and discussion.

Philosophy of Science, 87 (December 2020) pp. 1119–1129. 0031-8248/2020/8705-0029$10.00
Copyright 2020 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/710540 Published online by Cambridge University Press

This problem of understanding what meaning aggregate data have for individual patients has gone unresolved, despite the rise of evidence-based medicine (EBM) in recent decades. EBM provides principles and tools to assist in the evaluation and use of epidemiological evidence in individual patient care. Among these tools are evidence-based decision aids like the graphic provided by the Mayo Clinic to help physicians and patients understand the results of epidemiological evidence on statins and heart attacks (fig. 1). Decision aids are now available for all sorts of interventions and medical conditions, from screening for cancer to antidepressants for depression.

The decision aid in figure 1 represents the risk of heart attack in a certain low-risk population, as well as the effect of a statin, a drug for treating cholesterol, on lowering risk. It is derived from clinical trials studying the effectiveness of statins. The bottom line: if a certain population took a statin for 10 years, there would be one fewer heart attack per 100 people compared to that same population if untreated (a counterfactual comparison).

If you are an individual patient (or their clinician), you ultimately want to know whether the statin will benefit you (or your patient). Are you a smiling dark gray person, a frowning light gray person, or a brimming jet black person? This question is precisely not one that the evidence and the decision aid answers. A preferred measure of effect size in EBM is the number needed to treat (Guyatt et al. 2015), that is, the number of individuals a physician would need to treat with the intervention to prevent one outcome. The number needed to treat represents the same kind of information captured by decision aids like figure 1.

EBM tries to bridge the divide between populations and individuals using a unique concept of medical risk, the 'patient's risk', an individual probability. The patient's risk has become an important target of prognostication and

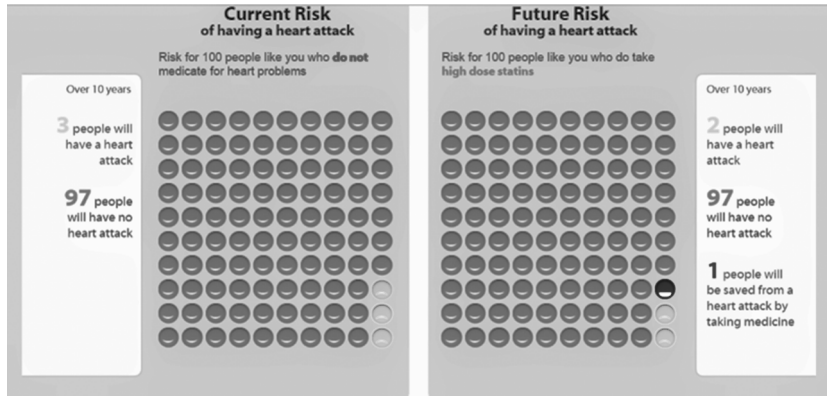


Figure 1. Decision aid representing the effect of a statin on risk of heart attack (Mayo Clinic 2017). Color version available as an online enhancement.

https://doi.org/10.1086/710540 Published online by Cambridge University Press

treatment in health care (Aronowitz 2015). Clinicians often speak of predicting and managing risk. Indeed, preventive interventions like statins are understood as 'lowering' or 'reducing' one's risk of some undesirable outcome. The field of risk communication sprang up to study empirically how people (especially patients and physicians) understand risk and to enhance discussions through tools like graphical aids (the psychology of risk). Yet, the meaning or interpretation of the patient's risk as a probability remains virtually unexplored. Goodman, an epidemiologist, writes: "The concept of risk has undoubtedly become an enormously useful and essential part of medical science, but evidence-based medicine must grapple with the fact that in the individual patient, one of its pillars—the concept of probability—is ambiguous and elusive" (1999, 606). Effective risk communication hinges on the problem of risk interpretation, with which philosophers can be of some help.

Here, I show how EBM's view of evidence and risk makes the problem of the meaning of population evidence difficult to resolve, and I propose an epistemic interpretation of individual patient probabilities as a corrective. In section 2, I illustrate EBM's view of epidemiological evidence qua medical evidence as measuring the patient's risk or the change in the patient's risk. I argue that the concept of the patient's risk is interpreted ontically as a propensity and that it is problematic in several ways. In section 3, I contrast the ontic interpretation with an epistemic (re)interpretation of individual patient probabilities as credences, in which epidemiological evidence is instead seen as informing medical uncertainty.

2. The Patient's Risk as an Ontic Probability. After evaluating the quality of epidemiological evidence (e.g., the likelihood of bias), the next step in the EBM process is to ask, what are the results (Guyatt et al. 2015)? In prognosis, a measure called the absolute risk (AR), the relative frequency of the outcome, is regarded as perhaps the most important information a study can provide. In therapy, some measure of effect size like the absolute risk reduction (ARR), the difference in AR between study groups, is thought to be the most useful takeaway.

On their own, these measures do not tell us how justified we are in believing that an individual patient will have a certain outcome or that the treatment will prevent it. Think back to the left-hand side of figure 1. Once we have summarized the AR—3% of people will have a heart attack—what should we infer about a given patient, about whether he or she will have a heart attack? In EBM, the fact we infer is about another kind of risk, the *patient's risk*. Prima facie, the patient's risk is meant to connect ARs and ARRs to the patient. The American College of Cardiology evidence-based guidelines for cardiovascular disease prediction says: "It is important to note that *risk* estimation is based on group averages, which are then applied to *individual* patients in practice" (Goff et al. 2014, S53; my emphasis). According

https://doi.org/10.1086/710540 Published online by Cambridge University Press

to their guidelines for cholesterol treatment: "statin therapy is recommended for *individuals* at increased ASCVD [cardiovascular disease] risk who are most likely to experience a net benefit in terms of the potential for ASCVD *risk reduction*" (Stone et al. 2014, S7; my emphasis). Not all experts always agreed with this approach (recall Hill's skepticism); nonetheless, it is the dominant approach in health care today.

There is good reason to suspect that conceptually the patient's risk is different from the AR, the finite frequency in the population. For one, the patient's risk is about an individual, not a population. Moreover, we can alter the patient's risk by providing the patient (as opposed to the whole population) with a treatment like a statin drug. In fact, numerous authorities in EBM and epidemiology understand the patient's risk as a probability (Goodman 1999; Djulbegovic, Hozo, and Greenland 2011; Guyatt et al. 2015). Probabilities are typically considered to be closely linked to finite frequencies like the AR yet distinct. The value of the patient's risk is taken to equal the AR, and the value of the change in the patient's risk due to treatment is taken to equal the ARR or some other AR-derived measure of the effect size. However, the patient's risk is inferred from the AR, and the change in the patient's risk due to treatment is inferred from the ARR.

In Fuller and Flores (2015), we call the standard model used in EBM to make these inferences the Risk Generalization-Particularization Model. We argue that the inferences rely on nontrivial and sometimes problematic assumptions. One of these assumptions is that the target population is a chance setup, like a lottery. We can then model the population using a hypothetical urn containing colored balls. Imagine that the treated and untreated diagrams in figure 1 represent two such urns. When a patient presents for risk assessment and treatment, it is analogous to choosing one of the two urns, randomly drawing a ball from it and then waiting 10 years to find out what its identity will be: dark gray (smiling) or light gray (frowning). The single case probability that the drawn ball (patient) will be light gray (have a heart attack) is given by the finite frequency of balls—in figure 1, the counterfactual probabilities are 3% (untreated) and 2% (treated). In the next section, I argue that we should interpret these probabilities epistemically. But how does EBM interpret 'the patient's risk', which is usually equated with the probability provided by our urn model?

How do we interpret *any* probability? There are several ways of distinguishing the kinds of interpretations on offer, including subjective versus objective interpretations (Gillies 2000) and belief-type versus frequency-type interpretations (Hacking 2001). Both of these dichotomies separate epistemic conceptions of probability (where probability refers to an epistemic notion like degrees of belief or the bearing of evidence on a hypothesis) from ontic conceptions of probability (where probability refers to something independent of our beliefs or evidence like frequencies or propensities).

https://doi.org/10.1086/710540 Published online by Cambridge University Press

I first consider whether EBM's concept of the patient's risk is interpreted epistemically in medicine. On this view, the patient's risk might be a certain degree of belief the physician has (or should have) that the patient will develop the outcome. Or it might be the logical support that the proposition 'the patient will develop the outcome' receives. However, a problem arises when we try to understand what it means to say that a statin lowers the patient's risk of heart attack. Seemingly, it means that the treatment causes the risk to be lower. On any realist theory of causation, causes exert a real effect in the world outside our minds. Sure, my belief that you will have a heart attack might be diminished if you take the medication, but the effect to which we are referring when we say that the intervention lowers your risk is not the effect the intervention has on my beliefs.

The patient's risk is more likely an ontic concept, locating probabilities in the world. One such interpretation to consider is frequentism, in which probabilities are either finite or 'long-run' frequencies. Yet the finite frequency just is the AR, so interpreted this way the patient's risk would fail to make population data any more meaningful for the individual. Meanwhile, the long-run or limiting frequency would be the proportion of light gray (frowning) balls drawn from the urn if we drew a ball randomly and then replaced it ad infinitum. That interpretation would only expand the population and our problem. Either variety of frequentism is prone to the notorious difficulty (if not outright futility) of making sense of a frequentist probability for a single case and, thus, is not a good candidate for EBM's singularist concept of the patient's risk.

The most plausible classic ontic interpretation of the patient's risk is as a propensity, a physical property of the single case that tends toward the outcome. Consistent with the way guidelines and clinicians speak about the patient's risk, it is best conceived as the individual's propensity toward some clinical outcome, and a lowering of the patient's risk due to an intervention is best conceived as a change in that propensity. The value of the patient's risk would then represent the strength of that propensity or the frequency of the outcome that the propensity manifests. Many philosophers are quite comfortable ascribing propensities to radioactive atoms or coin tosses, but are clinical outcomes genuinely chancy in this way?

Clinical outcomes result from biomedical mechanisms. For example, heart attacks result from coronary atherosclerosis, the buildup of a fatty plaque in the innermost layer of the coronary arteries supplying the heart due to the accumulation of LDL-cholesterol. A heart attack usually results from plaque rupture, leading to a blood clot that cuts off blood flow to heart muscle and deprives it of oxygen. A statin is thought to work by lowering LDL-cholesterol (among other potential effects). To say that outcomes like heart attack and statin prevention are chancy is to suggest that pathogenic mechanisms like these are chancy, which is far from accepted or obvious. Thus, the first difficulty

https://doi.org/10.1086/710540 Published online by Cambridge University Press

for the propensity interpretation is that the existence of propensities for clinical outcomes (never mind macrolevel outcomes generally) is not a given.

Viewing epidemiological evidence as measuring propensities invites further problems. First, the aggregate propensity or the change in aggregate propensity in an epidemiological population may be an unreliable estimate of individual propensities or changes in propensity, respectively. As a measure of propensity, the AR would first and foremost be a measure of aggregate or average propensity in a population. A propensity theory for medicine would have to assume that different patients have different propensities because ARs generally vary by clinical subpopulation according to risk factors. The question is whether the aggregate propensity is a good estimate of the individual propensity for most individuals. Similarly, the ARR would be a measure of the aggregate reduction in propensity in a population. As the individual treatment effect can vary among members of the population (Kravitz, Duan, and Braslow 2004), we should wonder whether the aggregate change in propensity is a good estimate for most individuals.

Whether an aggregate propensity (or change in propensity) is a good estimate for most individuals depends on the distribution of individual propensities (or changes in propensity) in the population. If the distribution is narrow and unimodal, then the true value for most individuals will cluster around the mean, so that the aggregate will usually be a pretty good estimate for the individual. If the distribution is not narrow or not unimodal, then many individuals will cluster away from the mean, and the aggregate will often be a poor estimate for the individual. As an illustration, if the clinical outcome is fully determined rather than governed by propensities, then individual probabilities and reductions in probability will only have two values, 0% or 100%, which means that the AR or ARR is aggregating an extreme bimodal distribution. What kind of distribution typically occurs in medicine?

It is difficult to know because virtually the only information we have about risks comes from aggregate studies. We do know (e.g.) that the AR of heart attack or stroke will vary widely from less than 1% to over 50% in common subgroups of patients (Goff et al. 2014), which suggests a large spread of cardiovascular propensities in the population. We have tools to further stratify patient risk, but not down to the level of the individual. In general, we should worry that any AR and ARR might be a poor estimate for the individual unless we have evidence to suggest otherwise.

One reason to treat this possibility as worrisome is that assuming that the aggregate is a good estimate for all individuals has two alarming consequences if we are mistaken. First, we ascribe a definite nonzero individual risk to everyone. Because individual risks are increasingly treated as diseases in their own right (Aronowitz 2015), this ascription turns everyone into a patient (e.g., a ‘cardiovascular risk patient’) with an individual risk that might need monitoring and intervention. Moreover, we assign a nonzero

https://doi.org/10.1086/710540 Published online by Cambridge University Press

potential individual risk reduction to everyone, which makes it appear as though the intervention is universally effective—everyone stands to benefit by having his or her own risk lowered. The assumption of an equal distribution of individual risk and risk reduction could lead to high health care anxiety, high health care utilization and costs, and overtreatment.

The final problem with EBM's view of epidemiological evidence as measuring individual risks as propensities is that individual propensities are just not what epidemiological studies directly measure, even if it turns out that some of the time an epidemiological study's results accidentally provide a good approximation of individual propensities. A 3% individual propensity toward heart attack is a property of the individual, analogous to a 100-faced die that has a 3% propensity toward rolling light gray because three of its 100 faces are light gray. We infer that there is a 3% probability of rolling light gray by studying the die's properties (number of faces, weighting, symmetry). In contrast, we modeled our epidemiological study results as an urn containing balls. We infer that there is a 3% probability that the ball we have drawn is light gray not from facts about this ball but from the frequency of light gray balls in the urn. How then should we interpret this individual probability?

In summary, the patient's risk is interpreted ontically in EBM, as a propensity. I am not denying that there might be clinical propensities, and the alternative I will now propose is agnostic to their existence. However, EBM's concept of the patient's risk faces serious problems: the existence of clinical propensities is not a given, aggregate epidemiological outcomes might be a poor estimate of them (which could lead to overdiagnosis of individual risk and overtreatment), and they are not what epidemiological evidence directly measures. Rather than individual propensities ('the patient's risk'), EBM should reconceive of the role of epidemiological evidence as supplying *epistemic* patient probabilities.

3. An Epistemic Reinterpretation. To provide a head-to-head comparison between EBM's interpretation of individual patient probabilities and the interpretation I endorse, I follow EBM in assuming that epidemiological studies measure ARs and effect size statistics derived from the AR like the ARR. I further follow EBM in inferring individual patient probabilities that are equal in value to the AR. I am not necessarily advocating this approach, but I will provide a better interpretation of these probabilities if we do choose to infer them.

The implicit interpretation of these probabilities in EBM is ontic: the patient's risk is something mind independent, most likely a propensity. Assuming you are a member of the untreated population represented in figure 1, your individual risk or propensity of having a heart attack in the next 10 years is 3%. The virtue of a statin is that it lowers your propensity of heart attack from 3% to 2%. In comparison, an epistemic interpretation of these probabilities

https://doi.org/10.1086/710540 Published online by Cambridge University Press

sees them as the probability of a hypothesis about you. The relevant prognostic hypothesis is *H*: you will have a heart attack in the next 10 years if untreated. The epidemiological evidence represented in figure 1 can be viewed as providing probabilistic support for this hypothesis: 3% represents the probability of *H* conditional on the evidence *E* ($p(H|E) = \text{AR}$).

There are various epistemic interpretations on the menu, including classical, logical, and credence varieties (Gillies 2000). I endorse interpreting patient probabilities as (rational) credences or degrees of belief over other epistemic interpretations for several reasons. First, credence is currently the leading contender among epistemic interpretations in philosophy, statistics, and medicine. Thinking about credences in prognosis and therapy might not be a huge conceptual leap for EBM, as Bayesian inference is commonplace in evidence-based diagnosis (Guyatt et al. 2015). Moreover, uncertainty is widely recognized in medicine as an important feature of clinical reasoning and decision making (Djulbegovic et al. 2011; Tonelli and Upshur 2019). Medical evidence is often seen as having some bearing on medical uncertainty, hopefully reducing it by providing new information but sometimes increasing it by challenging existing thinking. Credences are a ready and useful way of representing uncertainty. Epidemiological evidence informs uncertainty by grounding our credences in frequencies.

Finally, a credence interpretation is well suited to our probability model: an urn containing colored balls. We have drawn a ball randomly from the untreated urn in figure 1. The probability that we have drawn a light gray (frowning) ball is 3%, which can be interpreted as our credence that the ball is light gray given evidence of the frequency of light gray balls in the urn. I do not mean that this credence is the one that a particular physician will have, descriptively speaking. I mean that it is the rational credence we ought to adopt if the epidemiological evidence is our evidence. Asserting *E* as our evidence makes several assumptions beyond the important one that we have no independent evidence (these assumptions are explored in Fuller and Flores [2015]).[1] It amounts to claiming that our probability model represents our patient and our patient population well enough. If we grant these assumptions, the justification for setting our credence to the AR can be provided by one of several rational principles, including the Principal Principle (Lewis 1980) or the Principle of Direct Probability (Hacking 2001). I do not have space to discuss these principles here.

1. If we do have 'independent evidence' (evidence not bearing on *E* itself), our epidemiological credence could still be regarded as rational if we treat it as our prior probability that we can then update using the independent evidence. We might also have 'nonindependent evidence' (bearing on *E* itself), supporting the assumptions on which *E* relies, for instance, the assumption that the clinical trial population is representative of the target population.

https://doi.org/10.1086/710540 Published online by Cambridge University Press

In section 2, I raised the concern that because their populations are heterogeneous, epidemiological studies might provide an inaccurate estimate of the patient's true individual propensity. Our rational credence cannot be inaccurate in the same way as the propensity estimate because it is not intended as some ideal credence equal to the patient's propensity. It rather represents our uncertainty given variability in the outcome in the population, captured by the aggregate outcome. Some of this variability may be due to stochasticity: the outcome might actually be governed by propensities. However, some of the variability also results from causal heterogeneity. The study's aggregate outcome packages together any genuine stochasticity with causal heterogeneity. Its weakness in estimating an individual propensity is simultaneously its strength in serving to represent the physician's uncertainty to the extent that the physician does not know where in the propensity distribution the particular patient is located.

In section 2, I rejected an epistemic interpretation of the patient's risk because it is unfaithful to how the concept is used in practice. Physicians speak of lowering the patient's risk with an intervention, which is best understood as a causal effect of the intervention on the patient rather than its effect on our beliefs. Similarly, the ARR and other measures of effect size present a challenge for my epistemic alternative to the patient's risk. How can our credences reflect the causal import of the ARR, which measures not just a difference in frequencies but a difference caused by the intervention?

One tantalizing possibility is suggested by the depiction of the brimming jet black person in figure 1. The decision aid assumes that the ARR of 1% reveals the frequency of individuals for whom the statin prevented a heart attack. We might then infer the probability that this patient—presenting randomly from the population—will have a heart attack prevented by a statin is 1%. This probability could be interpreted as our credence in the hypothesis that a statin will prevent a heart attack for the patient.

Unfortunately, the decision aid goes beyond the trial evidence in ways that might not be justified. The trials simply determined that a statin reduced the frequency of heart attack by 1% on net. It could be that the statin prevented a heart attack in 1%, but it could instead be (e.g.) that a statin caused a heart attack in 2% while preventing one in 3%. These sorts of opposite or 'paradoxical' effects sometimes occur in medicine, like when an anticoagulant prevents death in one person by preventing stroke but causes death in another by producing a major gastrointestinal bleed or when an antibiotic prevents death in one by treating a life-threatening infection but causes death in another by triggering anaphylaxis. Furthermore, there may be nobody in the trial for whom the statin prevented a heart attack in the deterministic counterfactual sense of causing the absence of a heart attack when one otherwise would have occurred. It could be that a statin lowered the chance of a heart attack in one or more individuals who otherwise would have had a chance of heart

https://doi.org/10.1086/710540 Published online by Cambridge University Press

attack less than one. If we do not wish to rule out the possibility of paradoxical effects or chancy causation, we must depart from the decision aid's belief in the one jet black person.

Thankfully, we do not need this additional assumption to make clinical trial evidence meaningful on an epistemic interpretation. Our urn model provides the answer. The two urns in figure 1 generate two epistemic probabilities: cr($H|E$&untreated) and cr($H|E$&treated). Some of the epidemiological evidence in E comes from cohort studies that followed populations of patients over time and measured the untreated AR: these studies are evidence for the frequency in the untreated urn. However, some of the evidence in E comes from clinical trials that measured the effect size (e.g., the ARR). The effect size allows us to predict the frequency in the treated urn from the frequency in the untreated urn because it quantifies the difference that the intervention makes in the population. Thus, our credences reflect relevant causal facts because they are conditional partly on causal evidence (from clinical trials). We can thus make a rational decision about whether to use the intervention partly by comparing our treated and untreated credences.

An epistemic interpretation provides at least a partial solution to the problem of the meaning of population evidence for the individual. It does so by providing a meaning or interpretation of the individual probabilities that epidemiological evidence supplies: they represent our rational credence conditional on the evidence (and on several assumptions). Rather than measuring individual propensities (the patient's risk), the role of epidemiological evidence qua medical evidence is to inform medical uncertainty. The greatest benefit of the epistemic interpretation may be that it avoids the problems of the propensity interpretation discussed in section 2.

The epistemic interpretation has the additional benefit of restoring the importance of clinical judgment and expertise as well as the epistemic function of evidence to our understanding of epidemiological evidence in medicine. In its quest for objectivity, EBM has divorced epidemiological evidence from clinical expertise and from epistemic warrant: epidemiological evidence measures facts about individuals rather than providing epistemic support for hypotheses. Yet, patients go to their physician in part for the physician's medical opinion, his or her expert probability assignment. This epistemic probability need not be baseless; it should be grounded in objective facts supplied by evidence to allow for rational decision making. Similarly, we need not dispense with decision aids, which are often a helpful way of representing the evidence on which our credence is based. However, the evidence and the decision aids require interpretation to make them meaningful for the individual, and here the language of uncertainty should replace talk of the patient's risk.

Yet part of the problem of the meaning of population evidence for the individual remains unresolved. Our evidence-based credence is set to the

https://doi.org/10.1086/710540 Published online by Cambridge University Press

AR—a fact about the population—rather than reflecting facts particular to the person. The more relevant facts the population shares with a particular person, the more relevant the population evidence will be for that person. However, regardless of how relevant some population evidence is, my epistemic proposal at least makes this evidence interpretable for "a science of uncertainty and an art of probability."[2]

REFERENCES

Aronowitz, Robert. 2015. *Risky Medicine: Our Quest to Cure Fear and Uncertainty*. Chicago: University of Chicago Press.

Djulbegovic, Benjamin, Iztok Hozo, and Sander Greenland. 2011. "Uncertainty in Clinical Medicine." In *Handbook of the Philosophy of Science*, vol. 16, *Philosophy of Medicine*, ed. D. M. Gabbay, P. Thagard, and J. Woods. Amsterdam: Elsevier.

Fuller, J., and L. J. Flores. 2015. "The Risk GP Model: The Standard Model of Prediction in Medicine." *Studies in History and Philosophy of Biological and Biomedical Sciences* 54:49–61.

Gillies, Donald. 2000. *Philosophical Theories of Probability*. London: Routledge.

Goff, D. C., et al. 2014. "2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines." *Circulation* 129:S49–S73.

Goodman, S. N. 1999. "Probability at the Bedside: The Knowing of Chances or the Chances of Knowing?" *Annals of Internal Medicine* 130 (7): 604–6.

Guyatt, Gordon, Drummond Rennie, Maureen O. Meade, and Deborah J. Cook. 2015. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 3rd ed. New York: McGraw-Hill Education.

Hacking, Ian. 2001. *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press.

Hill, Austin Bradford. 1952. "The Clinical Trial." *New England Journal of Medicine* 247:113–19.

Kravitz, R. L., N. Duan, and J. Braslow. 2004. "Evidence-Based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages." *Milbank Quarterly* 82 (4): 661–87.

Lewis, David. 1980. "A Subjectivist's Guide to Objective Chance." In *Studies in Inductive Logic and Probability*, vol. 2, ed. Richard C. Jeffrey. Berkeley: University of California Press.

Matthews, J. Rosser. 1995. *Quantification and the Quest for Medical Certainty*. Princeton, NJ: Princeton University Press.

Mayo Clinic. 2017. "Statin Choice Decision Aid." https://statindecisionaid.mayoclinic.org/.

Stone, N. J., et al. 2014. "2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines." *Circulation* 129:S1–S45.

Tonelli, Mark, and Ross E. G. Upshur. 2019. "A Philosophical Approach to Addressing *Uncertainty* in Medical Education." *Academic Medicine* 94 (4): 507–11.

2. This expression, describing medicine, is attributed to the physician William Osler.

https://doi.org/10.1086/710540 Published online by Cambridge University Press