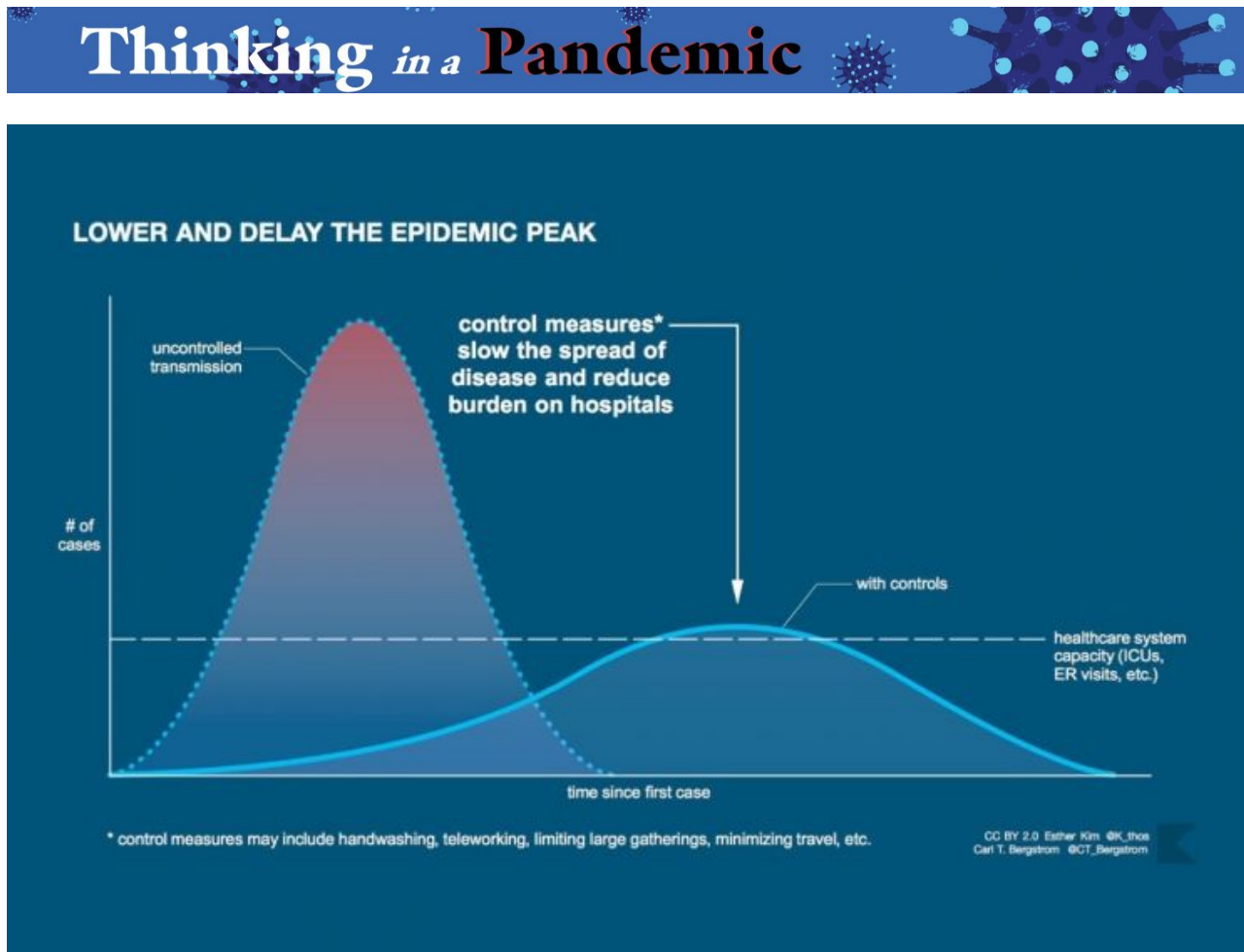


Models v. Evidence

[BR bostonreview.net/science-nature/jonathan-fuller-models-v-evidence](https://www.bostonreview.net/science-nature/jonathan-fuller-models-v-evidence)

Jonathan Fuller, *Boston Review*, May 5, 2020



Science & Nature

COVID-19 has revealed a contest between two competing philosophies of scientific knowledge. To manage the crisis, we must draw on both.

Jonathan Fuller

Image: Wikimedia Commons



The lasting icon of the COVID-19 pandemic will likely be the graphic associated with "flattening the curve." The image is now familiar: a skewed bell curve measuring coronavirus cases that towers above a horizontal line—the health system's capacity—

only to be flattened by an invisible force representing “non-pharmaceutical interventions” such as school closures, social distancing, and full-on lockdowns.

How do the coronavirus models generating these hypothetical curves square with the evidence? What roles do models and evidence play in a pandemic? Answering these questions requires reconciling two competing philosophies in the science of COVID-19.

To some extent, public health epidemiology and clinical epidemiology are distinct traditions in health care, competing philosophies of scientific knowledge.

In one camp are infectious disease epidemiologists, who work very closely with institutions of public health. They have used a multitude of models to create virtual worlds in which sim viruses wash over sim populations—sometimes unabated, sometimes held back by a virtual dam of social interventions. This deluge of simulated outcomes played a significant role in leading government actors to shut borders as well as doors to schools and businesses. But the hypothetical curves are smooth, while real-world data are rough. Some detractors have questioned whether we have good evidence for the assumptions the models rely on, and even the necessity of the dramatic steps taken to curb the pandemic. Among this camp are several clinical epidemiologists, who typically provide guidance for clinical practice—regarding, for example, the effectiveness of medical interventions—rather than public health.

The latter camp has won significant media attention in recent weeks. Bill Gates—whose foundation funds the research behind the most visible outbreak model in the United States, developed by the Institute for Health Metrics and Evaluation (IHME) at the University of Washington—worries that COVID-19 might be a “once-in-a-century pandemic.” A notable detractor from this view is Stanford’s John Ioannidis, a clinical epidemiologist, meta-researcher, and reliable skeptic who has openly wondered whether the coronavirus pandemic might rather be a “once-in-a-century evidence fiasco.” He argues that better data are needed to justify the drastic measures undertaken to contain the pandemic in the United States and elsewhere.

Ioannidis claims, in particular, that our data about the pandemic are unreliable, leading to exaggerated estimates of risk. He also points to a systematic review published in 2011 of the evidence regarding physical interventions that aim to reduce the spread of respiratory viruses, worrying that the available evidence is nonrandomized and prone to bias. (A systematic review specific to COVID-19 has now been published; it concurs that the quality of evidence is “low” to “very low” but nonetheless supports the use of quarantine and other public health measures.) According to Ioannidis, the current steps we are taking are “non-evidence-based.”

This talk of “biased evidence” and “evidence-based interventions” is characteristic of the evidence-based medicine (EBM) community, a close relative of clinical epidemiology. In a series of blog posts, for example, Tom Jefferson and Carl Heneghan of the Oxford Centre for Evidence-Based Medicine similarly lament the poor-quality data and evidence guiding action in the pandemic and even suggest that lockdown is the wrong call.

Models without evidence are blind, while evidence without models is inert.

In the other corner, Harvard's Marc Lipsitch, an infectious disease epidemiologist, agrees that we lack good data in many respects. Countering Ioannidis's hesitation, however, Lipsitch responds: "We know enough to act; indeed, there is an imperative to act strongly and swiftly." According to this argument, we could not afford to wait for better data when the consequences of delaying action are disastrous, and did have reason enough to act decisively.

Public health epidemiologists and clinical epidemiologists have overlapping methods and expertise; they all seek to improve health by studying populations. Yet to some extent, public health epidemiology and clinical epidemiology are distinct traditions in health care, competing philosophies of scientific knowledge. Public health epidemiology, including infectious disease epidemiology, tends to embrace theory and diversity of data; it is methodologically liberal and pragmatic. Clinical epidemiology, by contrast, tends to champion evidence and quality of data; it is comparatively more methodologically conservative and skeptical. (There is currently a movement in public health epidemiology that is in some ways closer to the clinical epidemiology philosophy, but I won't discuss it here.)

To be clear, these comparisons are fair only writ large; they describe disciplinary orthodoxy as a whole rather than the work of any given epidemiologist. Still, it is possible to discern two distinct philosophies in epidemiology, and both have something to offer in the coronavirus crisis over models and evidence. A deeper understanding of modeling and evidence is the key not only to reconciling these divergent scientific mindsets but also to resolving the crisis.



Models

Public health epidemiology uses theory, especially theory from other health sciences like microbiology, to model infection and understand patterns and causes of disease. Many of the epidemic models that the public and public health researchers alike have been voraciously consuming—including models produced by Imperial College London that informed the U.K. and U.S. coronavirus response—are SIR-type models. The theory underlying these models is old, originating in the Kermack-McKendrick theory in the 1920s and '30s, and even earlier in the germ theory in the second half of the nineteenth century. The SIR framework partitions a population into at least three groups: those who are susceptible to future infection (S), those who are currently infectious (I), and those who have been removed from the infectious group through recovery or death (R). An SIR model uses a system of differential equations to model the dynamics of the outbreak, the movement of individuals among the various groups over time.

The most important question we can ask of an outbreak model during a crisis is not whether its assumptions are accurate but instead how well it predicts the future—a

hard-nosed practical question rather than a theoretical one.

Other models in the SIR family add additional groups to these three basic ones, such as a group for those who are infected with the virus but not yet infectious to others. Agent-based models also represent infection dynamics (how the number of cases changes over time), but they do so by modeling behaviors for each member of the simulated population individually. Curve-fitting models like the one used by the IHME are less theoretical; they extrapolate from previous infection curves to make predictions about the future. All these different models have been used in the COVID-19 pandemic. The diversity of approaches, along with divergent estimates for model parameters, partly explains the range of predictions we have seen.

Public health epidemiology also relies on a diversity of data—from multiple regions, using a variety of methods—to answer any one scientific question. In the coronavirus pandemic, in particular, research groups have used estimates of multiple key parameters of the outbreak (infection rate, average duration of illness) derived from multiple settings (China, Italy) and produced by various kinds of studies (population-based, laboratory-based, clinically based) to make projections. Public health epidemiology is liberal in the sense of relying on multiple tools, including modeling techniques (the Imperial College team has used several models), and also in the sense of simulating various possibilities by tweaking a model's assumptions. Finally, its philosophy is pragmatic. It embraces theory, diversity of data, and modeling as a means to reaching a satisfactory decision, often in circumstances where the evidence is far from definitive but time or practical constraints get in the way of acquiring better evidence.

A formative scientific moment for the public health epidemiology tradition was the epidemiological research on smoking and lung cancer in the 1950s and '60s. Although lung cancer is not an infectious disease and SIR modeling played no starring role in this research, it featured a similar scientific approach and philosophical outlook. The public health epidemiology philosophy is especially necessary early on in an outbreak of a novel pathogen, when untested assumptions greatly outnumber data, yet predictions and decisions must still be made.

Neil Ferguson, one of the leading epidemiologists behind the Imperial College models, describes epidemic modeling as "building simplified representations of reality." The characterization is apt because SIR-type models have variables and equations meant to represent real features of the populations modeled. (Other types of scientific tools, such as black box neural nets used in machine learning, work differently: they do not attempt to mirror the world but simply to predict its behavior.) We could therefore ask how well an SIR-type model mirrors reality. However, the primary use of the models, especially early on in an epidemic, is to predict the future of the outbreak, rather than to help us explain or understand it. As a result, the most important question we can ask of an outbreak model during a crisis is not whether its assumptions are accurate but instead how well it predicts the future—a hard-nosed practical question rather than a theoretical one.

Public health epidemiology is pragmatic. It embraces theory, diversity of data, and modeling as a means of recommending policy.

Of course, predictive power is not totally unrelated to a model's representational accuracy. One way to improve the predictive prowess of model is to go out and collect data that can confirm or deny the accuracy of its assumptions. But that's not the only way. By running many simulations of the same model under different assumptions (so-called sensitivity analysis), a modeler can determine how sensitive the model's predictions are to changes in its assumptions. By learning from multiple different models, a scientist can also triangulate, so to speak, on a more robust prediction that is less susceptible to the faults of any one model. Both strategies were used in determining U.K. coronavirus policy.

Finally, often a single, more accurate prediction based on high-quality evidence is less useful than a range of modeling predictions that capture best-case and worst-case scenarios (such as the range of death counts the [White House coronavirus taskforce](#) presented at the end of March). It might be prudent to plan for the worst case and not only the most likely possibility. A pragmatic philosophy generally serves public health decision makers well.

However, when certain predictions based on plausible model assumptions would lead decision makers to radically different policy recommendations, the assumptions should be investigated with further evidence. A team at [Oxford University](#), for example, performed epidemic modeling specifically to illustrate that worrying coronavirus projections depend crucially on estimates of the number of individuals previously infected and now immune to the virus. It is this kind of uncertainty that serves as fodder for the evidence thumpers.



Evidence

Clinical epidemiologists are playing their own part in the pandemic: they are designing clinical trials of COVID-19 treatments, such as the World Health Organization–organized multi-country [Solidarity trial](#). In keeping with the high standards of evidence in the EBM movement, these trials are randomized: individuals are randomly allocated to receive one treatment or another (or a different combination of treatments). Although opinions on the exact virtues of randomization vary slightly, the most popular idea is that randomization reduces systematic bias. In a clinical trial, randomization eliminates selection bias, resulting in trial groups that are more representative or [comparable](#) in terms of causally relevant background features. Randomized studies are preferred because they can generate evidence that is less biased and more accurate.

The clinical epidemiology tradition cautions that theory can sometimes mislead us—for instance, by smuggling in unproven assumptions that have not been empirically established in human populations.

The concept of evidence is central to clinical epidemiology and EBM alike. Clinical epidemiology research produces evidence, while EBM experts critically appraise it. Good evidence, this tradition says, consists mainly in the results of clinical epidemiology studies. The tradition is generally suspicious of theory, including reasoning based on pathophysiology and models of disease. It often cautions that theory can sometimes mislead us—for instance, by smuggling in unproven assumptions that have not been empirically established in human populations. In the coronavirus case, models assume—based on experience with other pathogens, but not concrete evidence with the new coronavirus—that individuals who recover from infection will develop immunity against reinfection, at least in the short term.

A central concern for this philosophy is not the diversity but the quality of data. A founding principle of EBM is that the best medical decisions are those that are based on the best available evidence, and evidence is better if it consists of higher-quality data. EBM provides guidance on which evidence is best, but clinical epidemiological methods such as meta-analysis do not allow one to amalgamate diverse kinds of evidence. The tradition is also conservative in basing conclusions only on well-established empirical results rather than speculative modeling, preferring “gold standard” randomized studies to hypothetical simulations. Finally, this tradition is skeptical, challenging assumptions, authority, and dogma, always in search of study design flaws and quick to point out the limitations of research.

A formative moment for the clinical epidemiology tradition was the British Medical Research Council’s 1948 trial of streptomycin for tuberculosis, widely considered to be one of the first modern randomized clinical trials. This philosophy can be especially helpful as an outbreak of a novel pathogen evolves, as better evidence becomes available to scrutinize previous assumptions and settle unanswered questions. Clinical epidemiology has the expertise to contribute much of this evidence.

In advocating for evidence-based public health measures, Ioannidis suggests subjecting interventions like social distancing measures to randomized trials. His suggestion may not be feasible in the United States given multiple levels of governance over social distancing policies, among other logistical difficulties. But the suggestion that we should be studying the effectiveness of our public health interventions is as important as it is obvious, and clinical epidemiology is well placed to contribute to this endeavor. While public health epidemiology is adept at studying the distributions and determinants of disease, clinical epidemiology is at home in studying the effectiveness of healthcare interventions. (I do not mean to suggest that public health epidemiology lacks the resources to study its own interventions. Consider, for example, this clever impact study by Imperial College London.)

Measuring the effects of public health measures is far from trivial. Social distancing is not an intervention: it is a mixed bag of individual behaviors, some voluntary and some involuntary. These behaviors are represented in outbreak models by simulating reduced social interactions. The models sometimes suppose that certain specific interventions,

such as school or business closures, will produce particular patterns of social mixing. But the *effects* of specific interventions on patterns of social mixing is not the target of a classic SIR model. The modeler *inputs* patterns of social interaction; the model doesn't spit them out. (However, disease-behavior models *do* model social dynamics together with viral dynamics.) Rigorous research is needed to separate out the effects of individual interventions that have often been implemented simultaneously and are difficult to disentangle from independent behavior changes. Moreover, our interventions might have independent effects (on health, on the economy), and an outbreak model isn't broad enough in scope to predict these effects.

Institutionalized skepticism is important in science and policymaking. Too much of it is paralyzing, but it can provide a check on the pragmatic ethos of public health epidemiology.

Ioannidis also suggests a solution to the problem of inaccurate pandemic statistics: testing representative population samples, rather than relying on samples subject to sampling bias. In order to estimate the number of infected people and the growth of the pandemic over time, we can repeatedly sample from key demographics and perform diagnostic testing. Representative sampling and antibody assays can also help estimate the number of previously infected individuals who may be immune to reinfection. This information can help to rule out the Oxford scenario in which the susceptible population is much, much smaller than we think. It can also help in estimating the infection fatality ratio, the proportion of COVID-19 patients who die from their infection. Ioannidis argues that the infection fatality ratio has been greatly overestimated in certain contexts due to biased testing. Antibody testing has already begun in the United States and other countries, including a (not yet peer-reviewed) study by Ioannidis and colleagues estimating much higher prevalence of past COVID-19 infections in Santa Clara County than the official count. Ironically, the study was immediately criticized by scientists partly for its Facebook recruitment strategy on the grounds it may have resulted in a biased sample.

The key to proper representative sampling is clinical epidemiology's favorite motto: randomize it! Random sampling can overcome the sampling bias that has plagued modeling projections alongside the coronavirus. The clinical epidemiology tradition, transfixed with unbiased evidence, provides a ready solution to an urgent problem facing public health epidemiology.

The final gift that clinical epidemiology offers is its skeptical disposition. Institutionalized skepticism is important in science and policymaking. Too much of it is paralyzing, especially in contexts of information poverty that call for pragmatism—like at the outset of a pandemic involving a novel pathogen when we don't have gold-standard evidence to guide us, but inaction carries the risk of dire consequences. But clinical epidemiology's skeptical orientation can provide a check on the pragmatic ethos of public health epidemiology, preventing action from outrunning evidence, or at least helping evidence to catch up.

At the same time, a myopic focus on evidence alone would do a disservice to epidemiology. Were we to conduct randomized trials of public health interventions, the evidence generated would be inherently local—specific to the context in which the trials are run—because the effects of public health interventions (really, all interventions) depend on what other causal factors are in play. We can't simply extrapolate from one context to another. Similarly, we should not blindly extrapolate infection statistics from one location to another; all these parameters—the reproductive number, the attack rate, the infection fatality ratio—are context-sensitive. None of these statistics is an intrinsic property of the virus or our interventions; they emerge from the interaction among intervention, pathogen, population, and place.

It is theory, along with a reliance on a diverse range of data, that make coronavirus evidence collected in one place relevant to another. Evidence for the effects of interventions on social interactions must be combined with outbreak models representing those interactions. Evidence for age-stratified infection fatality ratios must be combined with local data about the age structure of a population to be of any use in predicting fatalities in that population. In an outbreak, models without evidence are blind, while evidence without models is inert.



Where does this clash of sensibilities leave us? In my own work, I have modeled prediction in evidence-based medicine as a chain of inferences. Each individual inference is a link forged from assumptions in need of evidence; the chain is broken if any assumption breaks down. In their book *Evidence-Based Policy* (2012), the philosopher of science Nancy Cartwright and the economist Jeremy Hardie represent predictions about the effectiveness of a policy using a pyramid. The top level, the hypothesis that the policy will work in some local context, rests on several assumptions, which rest on further assumptions, and so on. Without evidence for the assumptions, the entire structure falls.

We should welcome both of epidemiology's competing philosophies. Cooperation in society should be matched by cooperation across disciplinary divides.

Either picture is a good metaphor for the relationship between evidence and models. Evidence is needed to support modeling assumptions to generate predictions that are more precise and accurate. Evidence is also needed to rule out alternative assumptions, and thus alternative predictions. Models represent a multiverse of hypothetical futures. Evidence helps us predict which future will materialize directly by filling in its contours, and indirectly by scratching out other hypothetical worlds.

The need for evidence and modeling will not dissolve when the dust settles in our future world. In evaluating the choices we made and the effectiveness of our policies, we will need to predict what would have happened otherwise. Such a judgment involves comparing worlds: the actual world that materialized and some hypothetical world that did not. How many COVID-19 deaths did our social distancing measures prevent? We can

estimate the number of COVID-19 deaths in our actual socially distanced world by counting, but to predict the number of COVID-19 deaths in an unchosen world without social distancing we will need to dust off our models and evidence.

Just as we should embrace both models and evidence, we should welcome both of epidemiology's competing philosophies. This may sound like a boring conclusion, but in the coronavirus pandemic there is no glory, and there are no winners. Cooperation in society should be matched by cooperation across disciplinary divides. The normal process of scientific scrutiny and peer review has given way to a fast track from research offices to media headlines and policy panels. Yet the need for criticism from diverse minds remains.

I mentioned that the discovery that smoking causes lung cancer was a discipline-defining achievement for public health epidemiology, while the British Medical Research Council's streptomycin trial was a formative episode in the history of clinical epidemiology. The epidemiologist Austin Bradford Hill played a role in both scientific achievements. He promoted the clinical trial in medicine and also provided nine criteria ("Hill's Viewpoints") still used in public health epidemiology for making causal inferences from a diversity of data.

Like Hill, epidemiology should be of two minds. It must combine theory with evidence and make use of diverse data while demanding data of increasingly higher quality. It must be liberal in its reasoning but conservative in its conclusions, pragmatic in its decision making while remaining skeptical of its own science. It must be split-brained, acting with one hand while collecting more information with the other. Only by borrowing from both ways of thinking will we have the right mind for a pandemic.

While we have you...

...we need your help. Confronting the many challenges of COVID-19—from the medical to the economic, the social to the political—demands all the moral and deliberative clarity we can muster. In *Thinking in a Pandemic*, we've organized the latest arguments from doctors and epidemiologists, philosophers and economists, legal scholars and historians, activists and citizens, as they think not just through this moment but beyond it. While much remains uncertain, *Boston Review's* responsibility to public reason is sure. **That's why you'll never see a paywall or ads.** It also means that we rely on you, our readers, for support. If you like what you read here, **pledge your contribution to keep it free for everyone by making a tax-deductible donation.**

[Donate Today](#)