

How Wrong Were the Models and Why?

[A aier.org/article/how-wrong-were-the-models-and-why/](https://aier.org/article/how-wrong-were-the-models-and-why/)

April 23, 2020



Phillip W. Magness

– April 23, 2020 Reading Time: 8 minutes





The epidemiology models used to justify and extend the ongoing coronavirus lockdown are starting to come under much-needed scholarly scrutiny. A new working paper published by the National Bureau of Economic Research (NBER) presents a detailed statistical examination of several influential models, and particularly the study out of Imperial College-London (ICL) that famously predicted up to 2.2 million COVID-19 deaths in the United States under its most extreme scenario.

The ICL model presented an array of scenarios based on different policy responses, but this extreme projection – also referred to as its “do nothing” scenario – grabbed all the headlines back in March. Although the ICL paper described its own “do nothing” scenario as “unlikely” given that it assumed the virus’s spread in the absence of even modest policy and behavioral responses, its astronomical death toll projections were widely credited at the time with swaying several governments to adopt the harsh lockdown policies that we are now living under.

The Trump administration specifically cited ICL’s 2.2 million death projection on March 16th when it shifted course toward a stringent set of “social distancing” policies, which many states then used as a basis for shelter-in-place orders. In the United Kingdom, where the same model’s “do nothing” scenario projected over 500,000 deaths, the ICL team was directly credited for inducing Prime Minister Boris Johnson to shift course from a strategy of gradually building up “herd immunity” through a lighter touch policy approach to the lockdowns now in place.

Plainly, the ICL model shifted the policy responses of two leading world powers in dramatic ways.

Indeed, the ICL team played no small role in hyping the projections of its “do nothing” scenario, even as its own report downplayed the likelihood of that outcome in favor of more conservative projections associated with an array of social distancing policies and suspensions of public gatherings. On March 20th ICL lead author Neil Ferguson reported the 2.2 million death projection to the *New York Times*’s Nicholas Kristof as the “worst case” scenario. When Kristof queried him further for a “best case” scenario, Ferguson answered “About 1.1 million deaths” – a projection based on a modest mitigation strategy.*

It’s worth noting that even at the time of its March 16th public release, the conditions of the ICL’s “do nothing” scenario were already violated, rendering its assumptions invalid. Most governments had already started to “do something” by that point, whether it involved public information campaigns about hygiene and social distancing or event cancellations and the early stages of the lockdown, which began in earnest a week earlier. Voluntary behavioral adaptations also preceded government policies by several weeks, with a measurable uptick in hand-washing traceable to at least February and a dramatic decline in restaurant reservations during the first two weeks of March. When read in this context, Ferguson’s decision to hype the extreme death tolls of the “do nothing” scenario to the press in mid-to-late March comes across as irresponsible.

Nonetheless, the alarmist death toll projections dominated the public narrative at the time and – citing the ICL model – the United States went into lockdown.

A month later, it has become readily apparent that the 2.2 million death projection was off by several orders of magnitude, as was its UK counterpart of 500,000 projected fatalities. Ferguson and the ICL team shifted their public commentary to emphasize other scenarios with more conservative projections in the tens-of-thousands (in some cases this was misleadingly depicted as a revision to their model, although it actually used the milder scenarios in the original March 16th paper).

Nonetheless, the damage from the over-hyped ICL “do nothing” scenario was already done. Indeed, as of this writing, President Trump is still citing the 2.2 million projection in his daily press conferences as the underlying rationale for the lockdowns. The *New York Times*’s COVID reporter Donald McNeil was also still touting the same numbers as recently as April 18th, and even a month later it remains something of a social media taboo for non-epidemiologists to scrutinize the underlying statistical claims of credentialed experts such as Ferguson.

“Stay in your own lane,” we’re told, and let the experts do their own work. Epidemiology has its own proprietary methods and models, even as their most alarmist scenarios – the ones that Ferguson publicly hyped to the media a month ago – falter in visible and obvious ways.

Enter the new NBER paper, jointly authored by a team of health economists from Harvard University and MIT. Its authors conduct a measured and tactful scrutiny of the leading epidemiology forecasts, including the ICL model at the heart of the lockdown policy decisions back in March. Among their key findings:

“The most important and challenging heterogeneity in practice is that individual behavior varies over time. In particular, the spread of disease likely induces individuals to make private decisions to limit contacts with other people. Thus, estimates from scenarios that assume unchecked exponential spread of disease, such as the reported figures from the Imperial College model of 500,000 deaths in the UK and 2.2 million in the United States, do not correspond to the behavioral responses one expects in practice.”

As the authors explain, human behavior changes throughout the course of an epidemic. Even basic knowledge of the associated risks of infection induces people to take precautionary steps (think increased handwashing, or wearing a mask in public). Expectations about subsequent policy interventions themselves induce people to alter their behavior further – and continuously so. The cumulative effect is to reduce the reliability of epidemiological forecasts, and particularly those that do not account for behavioral changes.

If this sounds familiar, it is the critique that my colleague Will Luther made on March 18th, only two days after the ICL model came out. He similarly noted this implication when Ferguson shifted the emphasis of his public commentary to the more conservative scenarios in his model at the end of March. I also pointed to the importance of behavioral adaption around this time when considering the many policy responses to COVID-19, from public health advice to lockdowns to border checkpoints in certain states.

The NBER paper authors further critique the ICL paper and four other epidemiology models for overstating their own certainty about their many projection scenarios. Behavioral adaptation, among other factors, reduces the accuracy of long-term forecasting. The presentation of multiple scenarios also requires the adoption of a multitude of underlying assumptions about how these factors will play out given each policy choice made. Unfortunately, none of the epidemiology models they considered took sufficient steps to account for these complications.

The NBER study thus concludes:

“In sum, the language of these papers suggests a degree of certainty that is simply not justified. Even if the parameter values are representative of a wide range of cases within the context of the given model, none of these authors attempts to quantify uncertainty about the validity of their broader modeling choices.”

Epidemiological expertise may convey specialized knowledge about the nature of disease transmission that is specifically suited to forecasting a pandemic’s spread. But it does not exempt the modelers from social scientific best practices for testing the robustness of their

claims. Nor does it obviate basic rules of statistical analysis.

It would be a mistake to pit epidemiology as a field against its “outside” critics though, as the ongoing COVID-19 debates actually reveal a much more complex scientific discussion – including among medical experts and other specialists in pandemics. Around the same time the ICL model was released in March, distinguished medical statistician

John Ioannidis issued a strong warning for disease modelers to recognize the severe deficiencies in reliable data about COVID-19, including assumptions about its transmission and its essentially unknown fatality rates.

More recently, a team of epidemiologists based at the University of Sydney examined the performance of the influential Institute for Health Metrics and Evaluation (IHME) model out of the University of Washington at predicting next-day fatalities in each of the 50 states. Looking at daily results from March and early April, they concluded that as much as 70% of the actual daily fatality totals fell outside of the model’s 95% confidence interval, by either being too high or too low. This finding is not necessarily discrediting of the IHME researcher’s approach, but it does speak to the need for further refinements in their techniques while also cautioning against using its predictions as a basis for policy-making while uncertainty about its accuracy remains high.

As these examples reveal, epidemiology, health economics, and related fields that specialize in medical statistics are not a single “consensus” to be deferred to as a monolithic voice of expertise. Rather, they host necessary and sometimes sharply divided debates – including over COVID-19.

To illustrate the importance of statistical scrutiny, it helps to look to past epidemics and observe what similar debates tell us about the accuracy of competing epidemiological forecasts. In the late 1990s and early 2000s one such example played out in Great Britain concerning Creutzfeldt-Jakob Syndrome, better known by its common moniker of “Mad Cow Disease.”

In 2001 the *New York Times* ran a story on different epidemiological projections about the spread of Mad Cow Disease, highlighting two competing models.

The first model came from a team of Jerome Huillard d’Aignaux, Simon Cousens, and Peter Smith at the London School of Hygiene and Tropical Medicine (LSHTM). Using a variety of assumptions about the disease’s existing prevalence (some of them hotly contested) as well as observational data about the disease’s incidence prior to its highly publicized 1996 outbreak, the LSHTM model offered a variety of scenarios depicting an overall mild transmission pattern for the disease.

As Cousens told the *Times* in 2001, “No model came up with a number exceeding 10,000 deaths and most were far lower, in the range of a few thousand deaths” spread over the next decade. While the Mad Cow Disease literature continues to debate some of the underlying

assumptions of their model, the LSHTM team's mortality projections ended up fairly close to reality – at least compared to other models.

An estimated 177 people died from Mad Cow Disease in the UK in the wake of the 1996 outbreak. Disease mitigation measures persist in an ongoing effort to prevent a future outbreak from cattle-to-human transmissions including import/export restrictions on beef and the slaughter of cattle to contain the infection in livestock, but for the past two decades annual Mad Cow fatalities in humans have remained extremely rare.

When the 2001 *Times* story ran however, a different model dominated the headlines about the Mad Cow outbreak – one that projected a wide-scale pandemic leading to over 136,000 deaths in the UK. The British government relied on this competing model for its policy response, slaughtering an estimated 4 million cows in the process. The competing model did not stop at cattle either. In an additional study, they examined the disease's potential to run rampant among sheep. In the event of a lamb-to-human transmission, the modelers then offered a “worst case” scenario of 150,000 human deaths, which they hyped to a frenzied press at the time.

In the 2001 *Times* article, the lead author of this more alarmist projection responded to the comparatively tiny death toll projections from the LSHTM team. Such numbers, he insisted, were “unjustifiably optimistic.” He laid out a litany of problems with the LSHTM model, describing its assumptions about earlier Mad Cow Disease exposure as “extremely naïve” and suggesting that it missed widespread “underreporting of disease by farmers and veterinarians who did not understand what was happening to their animals.” He conceded at the time that he had “since revised [the 136,000 projection] only very slightly downward,” but expressed confidence it would prove much closer to the actual count.

The lead author of the extreme Mad Cow and Mad Lamb Disease fatality projections in the early 2000s is a familiar name for epidemiological modeling.

It was Neil Ferguson of the ICL team.

As with the present crisis, a high degree of uncertainty has loomed over epidemiological forecasts in the past. Such uncertainty is likely unavoidable, but it also produces a wide range of competing projections. When governments design policy based on epidemiological forecasts, their choice of the model to use could be the difference between a mild mitigation strategy and a large proactive intervention, such as the mass slaughter of livestock in the case of Mad Cow Disease or aggressive and wide-scale societal lockdowns in the case of COVID-19.

That choice, often made amid severe data limitations, is often presented to the public as an unfortunate but necessary action to forestall an apocalyptic scenario from playing out. But we must also consider the unseen harms incurred when politicians base decisions on a

modeled scenario that is not only unlikely but also wildly alarmist and likely exaggerated by the dual temptations of media attention and gaining the ear of politicians.

Given the high uncertainties revealed by statistical scrutiny of epidemiological models including among other medical experts, the presumption should go the other way instead. What is warranted is not bold political action in response to speculative models generated with little transparency and dubious suppositions, but rather extreme caution when relying on the very same models to determine policy.

**Correction. An earlier version identified the 1.1 million projection with the ICL “do nothing” scenario. It reflects a scenario with moderate set of mitigation policies.*

[READ MORE](#)

Phillip W. Magness

Phil Magness is a Senior Research Fellow at the American Institute for Economic Research.

He is the author of numerous works on economic history, taxation, economic inequality, the history of slavery, and education policy in the United States.



Books by Phillip W. Magness

Get notified of new articles from Phillip W. Magness and AIER.

[SUBSCRIBE](#)

Related Articles – Crisis, Economic Theory
