Harvard Data Science Review • Issue 3.1, Winter 2021

A Plague of Data

Theodore M. Porter¹

¹Department of History, Division of Social Sciences, University of California Los Angeles, Los Angeles, California, United States of America

Published on: Jan 29, 2021

DOI: <https://doi.org/10.1162/99608f92.6d743a66>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

The COVID-19 pandemic, more even than its predecessor epidemics, brought with it cascades of data. These were conceived as tools of numerical precision for the management of epidemic disease. Yet beyond inspiring a dismal awe, these numbers have failed in their mission to manage and ameliorate its effects. This failure is not to be blamed on the scientists, except perhaps in a secondary way. An effective intervention depended on a capacity to control human behavior. Epidemiological modeling, like so much of science, depends on a capacity to intervene effectively by altering behaviors on a massive scale. Indeed, behaviors changed, but in patchwork fashion, at least in most countries.

Sabina Leonelli ("[Data Science in Times of Pan\(dem\)ic](#)," this issue) here frames her reckoning in terms of alternative "imaginaries." This I think is correct, but the modest effectiveness that the situation seems to permit has depended on recognizing the constraints within which doctors, scientists, and expert managers were compelled to operate. 'Bureaucracy with imagination' would be an appropriate slogan. Their efforts were at times quite impressive, yet they were reined in on all sides. On this evidence, the data millennium is not yet upon us. New imaginaries of data and modeling, as Leonelli shows, are possible, but are unlikely to overcome an engrained economy or politics.

Nowhere did doctors and data scientists rule the roost. The American president, to begin with an especially deceitful case, was untainted by openness to real knowledge. Yet he was not straightforwardly dishonest, since he rarely made even the most perfunctory effort to deceive. He was canny enough to learn that the illness count could easily be reduced simply by reining in tests, a maneuver he advocated on national television. Pronouncements like these may, nevertheless, have contributed to public confusion about epidemic numbers. But the push for new data imaginaries faced some vexing obstacles. The most important data sources for this epidemic involved local, state, and national officials. In many places, including the United States, they were recorded locally, and were not made for easy data harmonization, still less to be comprehensively reimaged in the face of a new disease threat. The graphs and table published in press outlets like *The New York Times* showed dips during weekends and included piles of undated numbers that seemed to have been released in a block. Journalistic and scientific outlets tried to correct or at least provide indications of some irregularities, but in any cases of doubt were reluctant to take apparent liberties with the recorded data.

Other commentators were less constrained. Whole movements grew up to condemn official data (as well as election results, and much, much else), condemning explanations offered by the most plausible specialists. Even these did not always agree, and the water was muddied by people like Scott W. Atlas, the Republican-affiliated radiologist who claimed to know better than any epidemiologist and whose expert opinions consistently lined up with the administration. Even putting aside these dubious pronouncements, whole movements grew up to oppose the ostensible experts, while others justified their rejection of medical guidance in the name of freedom, the precious right to go maskless on beaches, in restaurants, and in groceries.

We seem, then, to face a battle between medical-scientific experts and politicized populists. Yet the obstacles to data-driven policy go further than this. The sorts of remedies that could stop or at least greatly reduce the infection rates were well-known almost from the beginning. China applied and enforced a strict policy of quarantine that apparently brought the infection under control. Other states, lacking the capacity or will to enforce such stern measures, faced hard choices about closing or opening schools, transportation systems, shops, hotels, and restaurants. Political leaders relied increasingly on rather loose standards to determine when restrictions should be stiffened or relaxed. On a collective level, these were broadly successful, though their validity, such as it was, never extended more than a few months into the future. Meanwhile, individuals made choices based on limited information. The most straightforward advice, to avoid contact as much as possible, and especially with those who might have been affected, was correct and obvious, though of course for many it was impractical or even impossible. Those who were not compelled to travel to schools or workplaces and who were not confined to prisons or other detention facilities could choose to limit contacts, avoid shops, wear masks, and so on, and in this way could improve their prospects of remaining healthy. It didn't require an epidemiologist to recognize this.

Some, including prominent political officials, put their faith in COVID-19 tests, and many discovered their limits of accuracy. The Rose Garden infestation, affecting officials and visitors right up to the president, provides memorable testimony to these limitations. These tests could have been made more reliable, but at the cost of much greater inconvenience and expense. They might have been good enough, especially when supplemented by close monitoring of contacts, if the disease level in the population could have been kept much lower.

Decisions as to how much risk a person cared to assume were available to the privileged, more or less in proportion to the degree of their privilege. These fortunate ones might have liked to know more about infection rates of flight attendants, grocery clerks, waiters, and so on, as a rough indication of the risks they would assume by entering such establishments. These possibilities raise questions of data ethics, which seem, however, to remain mostly hypothetical. The most effective use of data and tests was made by homes for the aged, whose infection rates diminished sharply after the first wave of infections due to testing and isolation. Such data partly inspired but mainly bolstered the expectations attached to these policies.

So long as infection rates remained low, data could be fairly effective, provided test results were effectively used. Sometimes they were not. On a national and international level, the fondest hopes of epidemiological data experts went unrealized, as, indeed, most seem to have anticipated. And this, I regret to say, is where the data imaginary properly leads us.

Disclosure Statement

Theodore M. Porter has no financial or non-financial disclosures to share for this article.

©2021 Theodore M. Porter. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.