Received: 25 March 2021 | Accepted: 25 March 2021
DOI: 10.1002/sim.8984

COMMENTARY

Statistics in Medicine WILEY

Rejoinder

Susan S. Ellenberg | **Jeffrey S. Morris**

Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence
Susan S. Ellenberg, University of Pennsylvania, 423 Guardian Drive, Philadelphia, PA 19104.
Email: sellenbe@pennmedicine.upenn.edu

Funding information
National Institute of Allergy and Infectious Diseases, Grant/Award Number: P30-AI045008

We very much appreciate the comments on our review, which have added much information and many insights into the response of the statistical community to the study of emerging infectious diseases and their impacts on public health. Many common themes appear in these commentaries, including the contribution of methods development in earlier outbreaks to modeling the COVID-19 pandemic, the importance of reliable data to support policy, and the vital need for statisticians to communicate effectively with the scientific, public health and political leadership, the media—and ultimately, the public.

The Drs Jewell provide more details on the problems of data collection and the consequent construction of reliable models to inform public health interventions in the current COVID-19 pandemic, contrasted to the situation with HIV when more comprehensive surveillance was performed. Of course, one contributing factor is the rapidity with which COVID-19 spread. AIDS was first recognized in 1981 but its threat to global public health was not immediately recognized. Nevertheless, one might certainly argue that our experience with HIV/AIDS, and to some extent with the 2009 appearance of H1N1 influenza, should have led more quickly to better systems of surveillance and data collection than we have achieved. This is an important national weakness that has impacted our health policy decision making as well as our ability to extract insights from our national health data. We agree that there have been missed opportunities in terms of national serology tests to track the pandemic over time and to assess durability of immunity after recovery; we have also missed opportunities to use our outstanding molecular biology capabilities to institute a broad viral sequencing strategy that would enable us to stay on the cutting edge of the emergence of SARS-CoV-2 variants.

They note the potential for assessing the COVID-19 death toll by considering year-to-year changes in all-cause mortality. We agree that excess death analyses are useful for measuring the aggregate impact of the pandemic, and benefits from its lack of dependence on accuracy of testing or death attribution, which is useful in a population with many skeptics who still doubt the seriousness of the pandemic. Additionally, by looking at state-by-state excess death analyses and assessing the correlation of time series of excess deaths with confirmed cases, it is possible to further infer what proportion of excess deaths are directly due to viral infection and not ancillary factors, which would be an interesting and important statistical contribution. A related approach is the calculation of life expectancy. Recent data have estimated that the life expectancy in the US decreased by a full year during 2020, a breathtaking statistic that demonstrates the undeniable impact of the pandemic.[1,2]

Jewell and Jewell also call attention to the lack of rigorous studies of nonpharmaceutical interventions. They note that important public health and societal questions, such as use of masks and whether schools needed to be closed, could have been subject to careful study but as yet have not been. A collaborative team at the University of Pennsylvania and Children's Hospital of Philadelphia is currently analyzing mask mandate effectiveness using county-level case counts for pairs of counties matched on factors known to affect viral spread, with one having mask mandates, the other not. A similar analysis focusing on the effect of school openings is also underway. We agree that such studies should be more prevalent and should be led by biostatisticians. The biostatistical community is perhaps less aggressive than it should be in seizing opportunities to rapidly initiate studies with high societal impact potential.

They note the problematic politicization of the epidemic response in the US. We wholeheartedly agree. The political divisiveness and distrust in the US has unfortunately compromised our ability to quickly establish surveillance programs, implement tried and true epidemic control measures such as contact tracing, or even agree upon basic mitigation strategies that could substantially suppress viral spread. The increasing tendency of individuals to seek information only from sources that reinforce their own viewpoints has exacerbated divisiveness and reduced the ability to find common ground. One might say the USA in the pandemic has been a sociological case study of dysfunctional divisiveness, and the poor outcomes that the country has experienced are evidence of some of its consequences.

They also comment on the importance of training scientists to communicate effectively to the broader community. Given that many misunderstandings about the pandemic have been driven by quantitative subtleties such as measurement error, missing data, selection bias, testing bias, confounding, and multiplicities, biostatisticians have an important role to play, not only in uncovering new insights from emerging data, but in communicating with the media and broader community. For example, one of us (J.M.) maintains a blog site (http://covid-datascience.com) that offers commentary on emerging information, debunks misinformation, and explains subtle quantitative issues that are important for the public to appreciate.

Dr Platt focuses on population-level validation of vaccine effectiveness and postapproval vaccine safety surveillance, an extremely important issue given the "warp speed" at which vaccines were developed and deployed. Vaccine hesitancy, an increasingly difficult public health problem, has been exacerbated by the speed with which COVID-19 vaccines have become available, raising concerns among many as to whether we know enough about the safety of these vaccines. Dr Platt notes the huge challenges ahead in the continued evaluation of the safety of the many vaccines now being used, involving both data collection and data interpretation, and notes established initiatives of the US and Canadian health systems that may contribute important information in this regard.

We have already seen analyses of vaccine efficacy based on electronic medical record (EMR) data coming out of Israel,[3-6] England,[7] and Scotland,[8] countries with national health-care systems with reliable electronic records systems. Such studies can provide rapid validation of the clinical trial efficacy results, particularly in patient subgroups. Some studies have also shed light on single-dose efficacy for vaccines studied as a two-dose regimen. In these observational studies, careful matching and/or other causal inference tools are necessary to avoid confounding that could bias assessments of vaccine efficacy; clear and transparent description of these methods are crucial. Statisticians need to be leading these analyses, ensuring that state-of-the-art methods are applied to these important questions.

Improved data on vaccine safety should help overcome vaccine hesitancy on the one hand, identify potential risks not seen in the clinical trials on the other. While the trials have so far demonstrated similar serious adverse event rates in vaccine and placebo groups, they were not large enough to detect rare serious adverse events that occur at rates of less than 1/5000, or provide precise estimates of adverse event rates in subpopulations not well represented in the studies (eg, immunocompromised, pregnant, or currently SARS-CoV-2 infected). Postapproval monitoring systems such as the Vaccine Adverse Event Reporting System (VAERS) in the USA include user-entered unverified information about adverse events appearing after vaccination. These data are difficult to interpret but can be useful for identifying "safety signals" that require more detailed study. Unfortunately, data from such systems, which are publicly available, can seriously mislead those who do not understand the large number of serious adverse events that will inevitably happen, unrelated to vaccine exposure, in the days or weeks after vaccination, given the large number of vaccinees.[9] The FDA and CDC also maintain systems using health system data[10] to investigate potential vaccine-related adverse events; while results of such investigations are often published, the process is not transparent to the public. Clear, transparent and rigorous population-level analyses of safety of vaccinated and matched nonvaccinated cohorts in countries with centralized EMR like Israel and the UK could help to reassure the public of vaccine safety, as well as identifying any potential vaccine-related safety issues.

DeGruttola and colleagues also address the need for better surveillance and remind us of the large cohort of HIV-infected individuals that have been followed for many years and which were crucial in defining the natural history of HIV infection. They describe a COVID-19 surveillance program in place at the University of California San Diego that could be a prototype of larger scale systems that could better define the natural history of infection and help clarify the longer term protection of vaccines and the emergence and behavior of new virus variants. They also aptly emphasize the importance of multidisciplinary team science in addressing the myriad challenges posed by pandemics driven by mysterious and dangerous pathogens. We agree with them that improved societal-level data collection and interdisciplinary team science are crucial to efficiently manage this crisis and prevent the next one; statisticians need to be key players in these efforts.

Dr Halloran, a world leader in the modeling of infectious disease, provides some history on epidemic modeling over the last several decades and more detail on the methods that have been developed to study transmission dynamics. While our focus was the comparison between statistical contributions to HIV/AIDS and COVID-19 research, Dr Halloran reminds us that the statistical work on epidemic modeling has been continuous through the decades—although the emergence of HIV/AIDS and most recently COVID-19 have certainly stimulated work in this area substantially. She notes that the statistical modeling and inference machinery, including nonlinear dynamical systems, causal inference with interference, multiscale mobility modeling, as well as innovative experimental designs crucial to these problems have come a long way, motivated by emerging infectious diseases. She highlights how the post-9/11 preparations for bioterrorism, the 2009 H1N1 pandemic, and the 2014-15 West African Ebola outbreak stimulated innovations in statistical modeling of infectious disease, and nicely complements the points we emphasized for HIV/AIDs and SARS-CoV-2/COVID-19.

Dr Brookmeyer calls attention to the “tip of the iceberg” phenomenon, seen in the earliest days of HIV/AIDS before the long latency period postinfection was recognized, and today when the presumably large number of infected but asymptomatic individuals contribute to the transmission of disease. This is undoubtedly a defining characteristic of SARS-CoV-2, and models that can account for these latencies are crucial for accurate projections. Along with other discussants, he calls attention to the need for statisticians to communicate effectively with public health leaders, policy makers and the public. Understanding the uncertainties in predictions arising from modeling, in particular the effects of modeling assumptions and sensitivity of findings to these assumptions, is critical to making optimal decisions about public health interventions and appropriately informing the public.

Dr Dean also emphasizes the importance of statisticians as scientific leaders and communicators and encourages statisticians to be proactive in outreach to the larger community via use of social media and by being open to requests from journalists interested in providing accurate information to the public. She very succinctly highlights the importance of our role, noting “We are able to bring nuance and insight to an often oversimplified public discussion, where an out of context point estimate can turn into a headline and then into a media frenzy.” This cannot be overemphasized—our skill set is needed to 1) elucidate the many nuances in the emerging data so that policy makers can make optimally informed decisions, and 2) provide the media with the information they need to accurately inform the public. Dr Dean has been a leader in this regard, not just by giving interviews, but by writing op-ed pieces and perspectives for national media such as the *Washington Post* and the *New York Times*; she also has a highly visible and impactful presence on Twitter, with over 100 000 followers. She has been a great role model for other statisticians in terms of engaging with the media, and through the media, with the public.

She also highlights the important methodological contributions of the HIV Vaccine Trials Network statisticians to the COVID-19 vaccine effort, and points out that the strategy of repurposing infrastructure put in place for earlier needs is essential during a fast-emerging pandemic. She affirms the primary role of statisticians in being advocates for data quality and accessibility, the limitations of which have contributed to difficulties in tracking (and hence managing) the pandemic, resulting in lost opportunities to learn more from the emerging data.

Dr Follmann notes the sense of urgency underlying the work of statisticians during both the HIV/AIDS and the COVID-19 pandemics, and in particular, the importance during the early days of AIDS clinical trials of engaging with the AIDS activists in order to better understand their concerns and to develop modified approaches to clinical trials that would address these concerns while still being able to yield reliable information about the effects of the treatments under study. There is no clear parallel to that initiative in the COVID-19 pandemic, but as Dr Dean stressed, effective communication with the population at risk—essentially, the entire public for COVID-19—is very important.

Dr Gail, whose seminal research with Dr Brookmeyer early in the HIV/AIDS era produced the method of back-calculation that allowed estimation of the latency period of HIV and consequently the number of individuals infected with HIV given the number diagnosed with AIDS, comments on the difficulty of estimating the size of the population infected with COVID-19. He notes, as do other discussants, the challenges posed by the varying and constantly changing societal measures to control spread, such as mask wearing and social distancing, and the challenges posed by the fast-moving and ever-changing pandemic on accurately tracking and understanding the data.

Dr Isham, as someone deeply involved in modeling the spread of HIV, notes our increased computational capacity in the current era of COVID-19, allowing more complex models to be implemented and thereby consideration of more factors in projecting the future course of the pandemic and estimating the potential impact of public health measures. Dr Isham emphasizes the importance of clear communication to policy makers and the public with regard to the degree of uncertainty around model predictions, and the assumptions that underlie these predictions. She also notes the inevitability of future pandemics, particularly with zoonotic origin, and urges the addition of social scientists and economists to the multidisciplinary group of statisticians, epidemiologists, computer scientists, applied mathematicians, biologists, virologists, and immunologists already heavily involved with pandemic modeling efforts. We strongly agree with her assessments. A broad multidisciplinary approach is necessary to understand the full range of impacts of the pandemic and potential responses so that policy makers can make fully informed decisions in the face of ongoing uncertainties and fast-accruing knowledge.

We find little, if anything, with which to take issue in the commentaries, and are grateful to all the authors for providing their insights and perspectives on what is the most challenging public health issue in any of our lifetimes.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Susan S. Ellenberg https://orcid.org/0000-0002-2364-5253

REFERENCES

1. https://www.cdc.gov/nchs/data/vsrr/VSRR10-508.pdf. Accessed March 25, 2021.
2. Andrasfay T, Goldman N. Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on the Black and Latino populations. *Proc Natl Acad Sci*. 2021;118(5):e2014746118. https://doi.org/10.1073/pnas.2014746118.
3. https://www.medrxiv.org/content/10.1101/2021.01.27.21250612v1. Accessed March 25, 2021.
4. https://www.medrxiv.org/content/10.1101/2021.02.08.21251325v1.full.pdf. Accessed March 25, 2021.
5. https://www.medrxiv.org/content/10.1101/2021.02.08.21251329v1. Accessed March 25, 2021.
6. https://www.medrxiv.org/content/10.1101/2021.02.06.21251283v1.full.pdf. Accessed March 25, 2021.
7. https://doi.org/10.2139/ssrn.3790399. Accessed March 25, 2021.
8. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3789264. Accessed March 25, 2021.
9. Ellenberg SS, Chen RT. The complicated task of monitoring vaccine safety. *Public Health Rep*. 1997;112:10-20.
10. https://www.cdc.gov/vaccinesafety/ensuringsafety/monitoring/vsd/index.html. Accessed March 25, 2021.

How to cite this article: Ellenberg SS, Morris JS. Rejoinder. *Statistics in Medicine*. 2021;40:2536–2539. https://doi.org/10.1002/sim.8984