

**COMMENTARY**

# Commentary on the role of statisticians in pandemics

**Ron Brookmeyer**

Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, California, USA

**Correspondence**

Ron Brookmeyer, Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, CA.

Email: rbrookmeyer@ucla.edu

I want to express my appreciation for the thoughtful article by Susan Ellenberg and John Morris that compares and contrasts aspects of HIV/AIDS and COVID-19 and the roles that statisticians are playing in addressing the challenges of these pandemics.<sup>1</sup> Public health emergencies require rapid response. Statistical contributions can profoundly inform public policy in health emergencies but only if the contributions are timely. The speed at which the COVID-19 crisis has unfolded has been staggering. The United States surpassed a cumulative COVID-19 death toll of 400 000 within 11 months of the first reported U.S. COVID case. In contrast, the HIV/AIDS cumulative death toll in the United States reached the 400 000 mark more than 16 years after the first AIDS reported case. Statisticians have risen to the enormous challenges presented by the COVID-19 pandemic and are making critical contributions on many fronts.

One area that statisticians are contributing is making sense of the myriad of models that try to forecast where the COVID-19 pandemic is heading. All too often COVID-19 model predictions that have appeared both in journals and the popular press are at great odds with each other.<sup>2</sup> Similarly, in the early days of HIV/AIDS, there were widely varying model predictions. For example, one of the first projections of the HIV/AIDS used the normal density to extrapolate AIDS incidence trends.<sup>3</sup> The extrapolation produced the anomalous prediction of rapidly decreasing AIDS cases despite the fact that all the data showed rising AIDS cases and concluded that the projected total size of the AIDS epidemic in the United States would be about 200 000 cases, a number that subsequently was discredited. The authors of that fallacious report cited William Farr and his work with smallpox in the 1830s for justification. Farr's curves for smallpox epidemics looked like normal curves and the normal density was used only by analogy. But of course, smallpox in the 1800s was neither AIDS in the 1990s nor COVID-19 in 2020. One of the earliest prediction models for COVID-19, also based on empirical extrapolation of curves, forecast that COVID-19 deaths in the United States would drop to near zero by summer 2020,<sup>4</sup> a prediction that has also been discredited.<sup>5</sup>

Lessons learned from both the COVID-19 and HIV/AIDS pandemics are that extrapolation of empirical trend curves, regardless how complicated the functional forms, can lead to very misleading results. Extrapolating COVID-19 death curves is especially perilous because of the relatively short time period between infection and death, and as such, death counts can take sharp abrupt turns just weeks after occurrence of events that are somewhat unpredictable such as community lockdowns, super-spreader events, and high travel days.

One parallel between COVID-19 and HIV/AIDS is that case surveillance data is only measuring the tip of the iceberg of the infected population. Statistical approaches are needed to estimate the size of the "hidden population" which refers to asymptomatic infected persons who are silently and unknowingly transmitting the infection to others. Back-calculation is a method that was developed to do that estimation for the HIV-infected population. Back-calculation performs statistical deconvolution to reconstruct historical HIV infection rates and thereby estimate the prevalence of HIV infection by using information about the incubation period of HIV and counts of AIDS.<sup>6</sup> Back-calculation relies on accurate estimates of the long and variable incubation period of HIV infection (the median incubation period in the absence of treatment is about 10 years). But, back-calculation methods have limited utility in the context of COVID-19 for several reasons. First, the incubation period of COVID-19 is much shorter (median about 5 days). Second, a significant but uncertain proportion of infected persons may never develop symptoms and thus these asymptomatic persons are not reflected at all in counts of cases, hospitalizations, or deaths. Third, the counts of incident COVID cases reported in surveillance data are more

incomplete and unreliable than counts of AIDS cases because of the scarcity of COVID-19 tests early in the pandemic and changing patterns of who shows up for tests.

COVID-19 modeling has appropriately focused on the transmission dynamics of infection. Transmission models such as susceptible-exposed-infected-recovered models rely on many assumptions such as how long people remain contagious, and rates of contact between persons, seasonality effects, and the introduction and effectiveness of infection control strategies such as physical distancing, wearing of masks or lockdowns. Agent-based modeling and simulations are important tools for implementing transmission models.<sup>7</sup> However, outputs from transmission models represented by smooth colorful curves can provide a false sense of precision and hide uncertainties. Builders of these models need to work especially hard to communicate both to the general public and policy makers the key assumptions that they relied upon. Unfortunately, no guidelines have been published to standardize the results of epidemic prediction models.<sup>8</sup> It is critically important that the statistical community advocate for transparency in transmission models. Standards for reporting results from transmission models need to be developed as has been done for clinical trials, observational studies, and systematic reviews.

Epidemics present special challenges in estimating time to event distributions (eg, the incubation period distribution). For example, the first study of the incubation period distribution of HIV infection was based on transfusion-associated AIDS cases and statistical methods were needed to address sampling biases in the data due to truncation. Challenges in estimating the incubation period of COVID-19 arise in part because the dates of infection are not known or not known precisely. Widely cited estimates of the COVID-19 incubation period distribution used statistical methods to address these issues.<sup>9,10</sup>

Pandemics present challenges in the design and conduct of therapeutic and prevention trials.<sup>11</sup> For example, in response to the HIV pandemic, statisticians reconsidered traditional approaches to therapeutic clinical trials and acknowledged that a growing epidemic affecting young people requires study design that address the public health urgency.<sup>12</sup> Surrogate markers were considered to shorten the duration of trials. Vaccine trials for highly infectious pathogens present critical design challenges because the size, geographic location, and duration of outbreaks are uncertain. Flexible and responsive vaccine trial designs are required as outbreaks ebb and surge in different locations.<sup>13,14</sup> Important considerations in designing vaccine trials include the trial target population, the randomization approach (eg, cluster vs individual randomization), and selection of trial endpoints (eg, infection, symptomatic disease, severe clinical disease or death.)

There will undoubtedly be future pandemics. When there is a new public health crisis, we do not have the luxury of time. To have public health impact, statisticians must rapidly become knowledgeable about different aspects of the pathogen, the natural history of disease, the completeness and accuracy of public health surveillance data, transmission dynamics of the infection, and therapeutic and prevention developments. Statisticians can contribute to a host of issues including prediction and epidemic modeling, complexities in estimating incubation periods and transmission rates, and challenges in designing and conducting trials. Statisticians can also contribute by engaging with the media to increase public understanding of statistical issues in predicting and responding to pandemics. While the statistical problems of current and future pandemics can be expected to vary, there is no doubt that statisticians and statistical science will continue to make impactful contributions that address the challenges of public health emergencies.

## DATA ACCESSIBILITY

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

## DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study

## ORCID

Ron Brookmeyer  <https://orcid.org/0000-0002-8639-3399>

## REFERENCES

1. Ellenberg S, Morris J. AIDS and COVID-19: a tale of two pandemics and the role of statisticians. *Stat Med*. 2021;40:2499-2510.
2. Brookmeyer R. Op Ed: predictions about where the coronavirus pandemic is going vary widely. can models be trusted? Los Angeles Times, April 22, 2020.
3. Bregman DJ, Langmuir AD. Farr's law applied to AIDS projections. *Jama*. 1990;263:1522-1525.

4. IHME Health Utilization Forecasting Team, Murray CJL. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. medRxiv; March 27, 2020:20043752 doi:<https://doi.org/10.1101/2020.03.27.20043752>.
5. Jewell NP, Lewnard JA, Jewell BL. Caution warranted: using the institute for health metrics and evaluation model for predicting the course of the COVID-19 pandemic. *Ann Int Med.* 2020;173(3):226-227.
6. Brookmeyer R, Gail MH. *AIDS Epidemiology: A Quantitative Approach*. Oxford, UK: Oxford University Press; 1994.
7. Halloran ME, Auranen K, Baird S, et al. Simulations for designing and interpreting intervention trials in infectious diseases. *BMC Med.* 2017;15(1):1-8.
8. Pollett S, Johansson M, Biggerstaff M, et al. Identification and evaluation of epidemic prediction and forecasting reporting guidelines: a systematic review and a call for action. *Epidemics.* 2020;33:100400.
9. Reich NG, Lessler J, Cummings DA, Brookmeyer R. Estimating incubation period distributions with coarse data. *Stat Med.* 2009;28(22):2769-2784.
10. Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Int Med.* 2020;172(9):577-582.
11. Institute of Medicine. *Methodological Challenges in Biomedical HIV Prevention Trials*. Washington, DC: The National Academies Press. 2008.
12. Byar DP, Schoenfeld DA, Green SB, et al. Design considerations for AIDS trials. *N Engl J Med.* 1990;323:1343-1348.
13. Dean NE, Gsell PS, Brookmeyer R, et al. Design of vaccine efficacy trials during public health emergencies. *Sci Transl Med.* 2019;11(499):eaat0360.
14. Dean NE, Gsell PS, Brookmeyer R, et al. Creating a framework for conducting randomized clinical trials during disease outbreaks. *N Engl J Med.* 2020;382:1366-1369.

**How to cite this article:** Brookmeyer R. Commentary on the role of statisticians in pandemics. *Statistics in Medicine.* 2021;40:2521–2523. <https://doi.org/10.1002/sim.8935>