

COMMENTARY

Comment on AIDS and COVID-19: A tale of two pandemics and the role of statisticians

M. Elizabeth Halloran^{1,2} 

¹Biostatistics, Bioinformatics, and Epidemiology Program, Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, Seattle, Washington

²Department of Biostatistics, University of Washington, Seattle, Washington

Correspondence

M. Elizabeth Halloran, Biostatistics, Bioinformatics, and Epidemiology Program, Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

Email: betz@fhcrc.org

Funding information

National Institute of Allergy and Infectious Diseases, Grant/Award Number: R37 AI032042

In the early 1980s, globally, just a handful of researchers worked at the interface of statistical analysis of infectious disease data and dynamic models. Reading Ellenberg and Morris on HIV and Covid-19, one might think that not much else had happened in the intervening 40 years. Since the beginning of the HIV pandemic, computational power, availability of data relevant to understanding natural history and dynamics of emerging infectious diseases, statistical methods for fitting non-linear dynamic models, whether stochastic or deterministic, and machine learning methods for prediction have all increased enormously.

A lot of research by statisticians and modelers on other infectious diseases besides HIV, including emerging infectious diseases, helped propel the fields of quantitative methods for infectious diseases forward from the 1980s onward. With the advent of the use of Markov chain Monte Carlo (MCMC) methods in statistics in the early 1990s, the ability to do estimation was much improved. Infectious diseases can be thought of as partially observed Markov processes. Many aspects of infectious disease transmission are not observed, such as infection times, clearance times, onset, and end of infectiousness, so latent period, incubation period, generation time, basic reproduction number, and other natural history parameters could be poorly estimated. MCMC methods improved the situation.^{1,2} Other statistical methods for inference for nonlinear dynamical systems in infectious disease have grown enormously.³ Causal inference used to make the no interference assumption, which usually does not hold in infectious diseases. Causal inference with interference started making headway in the early 1990s, but really took off well after 2000,⁴ fueled largely by social scientists as well.

After the 9/11 attack in 2001, the US government was interested in preparing for a bioterrorist smallpox attack and created a working group of modelers and statisticians under the Secretary for Health and Human Services. Preparation for a possible pandemic of pathogenic avian influenza was supported by the initiation of the Models of infectious Disease Agent Study (MIDAS) Network in 2004. Instead of avian influenza, the influenza A (H1N1) pandemic struck in early 2009. The H1N1 pandemic fueled more statistical and dynamic modeling advances, and data became increasingly available. Multiscale mobility modeling has made use of the enormous computational power to do global modeling,⁵ initially in the context of seasonal influenza, but with the ability to pivot to be used to other emerging infectious diseases. The West African Ebola outbreak 2014–2015, though not a global pandemic, challenged statisticians/modelers alike because of lack of data about the populations in Liberia, Guinea, and Sierra Leone. But it led to the novel ring vaccination trial in Guinea,⁶ which was the result of thinking at the interface of transmission, statistics, and cluster-randomized clinical trials.

After the West African Ebola outbreak came the Aedes-borne outbreaks of chikungunya and Zika in the western hemisphere. These also raised novel challenges for statistician/modelers. The interface of infectious disease models, statistical and epidemiological methods, and causal inference fields has grown tremendously in the past 40 years. All of this paved the way for a rapid response by the huge global community of statisticians, dynamic modelers, genome scientists, computer scientists, and epidemiologists when news of a novel coronavirus broke in early January 2020.

ORCID

M. Elizabeth Halloran  <https://orcid.org/0000-0002-3127-1757>

REFERENCES

1. O'Neill PD, Balding DJ, Becker NG, Eerola M, Mollison D. Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Appl Stat.* 2000;49:517-542.
2. Auranen K, Arjas E, Leino T, Takala AK. Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data. *J Am Stat Assoc.* 2000;95:1044-1053.
3. Ionides EL, Bretó C, King AA. Inference for nonlinear dynamical systems. *Proc Natl Acad Sci.* 2006;103(49):18438-18443. <https://doi.org/10.1073/pnas.0603181103>.
4. Hudgens MG, Halloran ME. Towards causal inference with interference. *J Am Stat Assoc.* 2008;103:832-842.
5. Balcan D, Colizza V, Goncalves B, Hud H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci.* 2009;106(51):21484-21489.
6. Ebola ça suffit ring vaccination trial consortium. The ring vaccination trial: a novel cluster randomised controlled trial design to evaluate vaccine efficacy and effectiveness during outbreaks, with special reference to Ebola. *BMJ.* 2015;h3740:351.

How to cite this article: Halloran ME. Comment on AIDS and COVID-19: A tale of two pandemics and the role of statisticians. *Statistics in Medicine.* 2021;40:2524–2525. <https://doi.org/10.1002/sim.8937>