

VIEWPOINT

Navigating Through Health Care Data Disrupted by the COVID-19 Pandemic

Kayoko Shioda, DVM, MPH

Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, Connecticut.

Daniel M. Weinberger, PhD

Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, Connecticut.

Makoto Mori, MD

Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, Connecticut; and Division of Cardiac Surgery, Department of Surgery, Yale School of Medicine, New Haven, Connecticut.

Corresponding

Author: Makoto Mori, MD, Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, One Church Street, Ste 200, New Haven, CT 06510 (makoto.mori@yale.edu).

The coronavirus disease 2019 (COVID-19) pandemic created abrupt changes in the health care system and health of the general population. The disruptions have influenced patient-, hospital-, and physician-level decisions and performance. For instance, the number of patients who sought health care for emergency cardiovascular conditions declined¹; outcomes of certain diseases, such as cancer, may have worsened²; and surgical procedures were reserved for patients with emergency conditions. Such changes are inevitably reflected in the health care data for future research. Analytic methods commonly used in clinical trials and observational studies, however, are often based on assumptions that these factors remain stable over time. Therefore, researchers and others interested in research should be cognizant of the unique characteristics of the data generated during the pandemic, and recognize that some studies may not produce reliable results. As some study designs are more likely to be affected, analyses should be appropriately adjusted to allow valid inferences to be made.

Pandemic-Related Disturbances in Health Care Data

The COVID-19 pandemic has influenced multiple levels of health care, including the composition of patient characteristics (eg, sociodemographic, disease severity, comorbidity), clinician-related factors (eg, center volume, patient load distribution among physicians, hospitals, and health systems), resource utilization (eg, intensive care unit admission), and clinical outcomes (eg, patient-reported outcomes, mortality, readmissions). Such temporary changes matter in ongoing and future research intending to inform patient care, because data generated during the pandemic may be markedly different for these key characteristics. For instance, certain demographic groups, such as women, African Americans, and Hispanics, disproportionately became unemployed and lost employment-based health insurance during the pandemic; the resulting perturbation of payer-based claims data is likely significant and uneven across sociodemographic strata.³

Unlike local disruptions of the data from natural disasters (eg, earthquakes, hurricanes), economic collapse (eg, the Greek government debt crisis starting in 2009), or health care worker strikes, the global scope of COVID-19 across diverse groups of people may create unique challenges to study designs that are unable to account for the local disruption by using unaffected age groups, regions, or countries as controls. In addition, the expected use of data sets combining the data generated before, during, and after the pandemic may obscure the populations to which the study findings apply.

Study Types That Could Be Biased by Changes Related to COVID-19

The changes in health care data are likely to affect various types of studies. For example, clinical predictive models, such as a model estimating the risk of 30-day postoperative mortality that was calibrated to the pre-pandemic data, may perform poorly in predicting the outcome for operations performed during the pandemic. Such models are unlikely to be calibrated to the increased background mortality rate during the pandemic, patient selection for triaged operations, and the increased risk of mortality associated with preexisting respiratory illness in patients with perioperative severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection.⁴ Furthermore, models fitted to data on operations performed during the pandemic might not be accurate in the postpandemic setting.

Although randomization may safeguard the data against such changes, randomized clinical studies still require cautious interpretation. Randomized studies that enroll participants during the pandemic may not be affected by pandemic-related disruptions because the disruption is equally distributed across the comparative arms. For the relationship between the intervention and outcome to be preserved, however, researchers should critically consider the possibility that the pandemic-related change interacts with the control and intervention arms differently. During the pandemic, certain interventions may be increasingly difficult to adhere to or perform as specified in the study protocol, such as chemotherapy sessions requiring multiple hospital visits or medication dose titration based on frequent blood draws. Multinational trials may underrepresent patients in severely affected countries, or enrollment may be skewed by the phase of the pandemic in a particular region. As a result, participants enrolled before, during, and after the pandemic may differ significantly from the population targeted for enrollment when the study was designed.

Studies that evaluate trends in disease incidence or outcomes over time are especially susceptible to bias. For example, vaccine evaluation studies comparing the disease incidence before and after vaccine introduction often assume that various temporal factors, such as health care access, health care utilization, and underlying health condition of the population, remain stable over time. This assumption may no longer hold during and after the pandemic, making conclusions drawn from time series analyses unreliable. In addition, social distancing measures have likely lowered transmission of not only SARS-CoV-2 but also other infectious diseases, such as influenza, and may exaggerate the measured effect of vaccines for other respiratory diseases. Furthermore,

lower vaccination rates during the pandemic could lead to a higher incidence of vaccine-preventable diseases long after the pandemic. This, in turn, could lead to biased estimates of the effects of a vaccine if these changes in vaccine uptake are not properly accounted for.⁵

Minimizing the Pandemic-Related Bias

Potential approaches to data disruptions during the COVID-19 pandemic include the following: (1) excluding the period of disruption; (2) using an analytical approach that can appropriately adjust for the disruption; and (3) relying on carefully designed randomized trials. Simply excluding the period of disruptions can be a challenge. Although the COVID-19 case numbers may roughly estimate this period, the duration of latent consequences for data and for the health care system likely exceeds a period defined by case counts. Identifying such a time frame is further complicated by the additional waves of COVID-19 infections in various parts of the world and the potential for the pandemic to become seasonal.⁶ Considering such a complex interplay, data-driven technique to identify the timing, magnitude, and pattern of changes (eg, sudden shift in mean, increase or decrease in the rate of change), may become increasingly important in postpandemic research. For example, change-point analysis detects shifts in the average number of cases or the rate of change and determines when these shifts occur.⁷

Randomized clinical trials that ensure that treatment and control arms are balanced over time would be least susceptible to

bias. Observational studies might use data on control conditions or other relevant variables to adjust for pandemic-related changes (eg, using instrumental variables analysis or synthetic controls). These methods are not immune to bias, and it is critical to ensure that the effects of the pandemic on the adjustment variables is understood.

Certain study types, however, may not produce reliable estimates. These include studies that rely on time-series data crossing the pandemic era and pre-post designs, including regression discontinuity or difference-in-difference analyses intending to evaluate the potential associations of a change unrelated to the pandemic (eg, guideline change in 2018). In such designs, accounting for pandemic-related changes via elaborate adjustment may render the interpretation too complex and conditional for the results to be broadly applicable.

Conclusions

Understanding the magnitude and the period of perturbation is essential for research relying on health care data generated during the COVID-19 pandemic. Although accounting for the pandemic-related changes may be feasible in some instances, the possibility that some study designs may be unable to produce reliable results should be acknowledged. It is important for consumers of such data, including the researchers, reviewers, journal editors, and readers, to recognize the uniqueness of the data generated during the pandemic and evaluate future studies accordingly.

ARTICLE INFORMATION

Published Online: October 12, 2020.
doi:10.1001/jamainternmed.2020.5542

Conflict of Interest Disclosures: Dr Weinberger reported receiving consulting fees from Pfizer, Merck, GlaxoSmithKline, and Affinivax and is principal investigator on a grant from Pfizer to Yale. These entities were not involved in any aspect of the current work. No other disclosures were provided.

REFERENCES

- De Rosa S, Spaccarotella C, Basso C, et al; Società Italiana di Cardiologia and the CCU Academy Investigators Group. Reduction of hospitalizations for myocardial infarction in Italy in the COVID-19 era. *Eur Heart J*. 2020;41(22):2083-2088.
- Sud A, Jones ME, Broggio J, et al. Collateral damage: the impact on outcomes from cancer surgery of the COVID-19 pandemic. *Ann Oncol*. 2020;31(8):1065-1074. doi:10.1016/j.annonc.2020.05.009
- Montenovo L, Jiang X, Rojas FL, et al. Determinants of disparities in Covid-19 job losses. National Bureau of Economic Research. Accessed August 2, 2020. <https://www.nber.org/papers/w27132>
- COVIDSurg Collaborative. Mortality and pulmonary complications in patients undergoing surgery with perioperative SARS-CoV-2 infection: an international cohort study. *Lancet*. 2020;396(10243):27-38. doi:10.1016/S0140-6736(20)31182-X
- Santoli JM, Lindley MC, DeSilva MB, et al. Effects of the COVID-19 pandemic on routine pediatric vaccine ordering and administration—United States, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(19):591-593. doi:10.15585/mmwr.mm6919e2
- Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*. 2020;368(6493):860-868. doi:10.1126/science.abb5793
- Kürüm E, Warren JL, Schuck-Paim C, et al. Bayesian model averaging with change points to assess the impact of vaccination and public health interventions. *Epidemiology*. 2017;28(6):889-897. doi:10.1097/EDE.0000000000000719