Guarding Against Seven Common Threats to the Credible Estimation of COVID-19 Policy Effects

Because there was no vaccine or effective drug treatment, nonpharmaceutical interventions (NPIs) such as social distancing, school closures, and large-scale population lockdowns were the centerpiece of the global response to the COVID-19 pandemic. NPIs aim to lower the basic reproduction number, $R_0(t)$, to below one by reducing the contact rate to prevent epidemic outbreaks. This can be achieved through at least three strategies: (1) suppression (i.e., severely restricting contacts in the population [e.g., through shelter-in-place policies]); (2) mitigation (i.e., flattening the transmission curve to enable the health care system to handle critically ill individuals at any time); or (3) herd immunity (i.e., letting the infection spread without active intervention to allow enough individuals to be infected and develop immunity to the disease, thus making its subsequent spread unlikely). A recent modeling study found that NPIs, particularly lockdowns, successfully reduced $R_0(t)$ across various European countries.[1] However, overly restrictive NPIs may impose substantial health, social, and economic costs.[2] Understanding which NPIs are most effective to prevent COVID-19 transmission is therefore critical to balance their public health benefits against these costs.

Although randomized control trials are impractical in the COVID-19 context, econometric techniques that carefully exploit the quasiexperimental nature of interventions can credibly estimate their causal effects. As more data emerge, these methods may enable researchers to assess whether various NPIs achieved their intended goals. Because different jurisdictions implemented different policies over time, a difference-in-difference (DiD) design can be used to estimate the causal effects of NPIs by comparing changes in COVID-19–related outcomes before and after the implementation of a given policy in a locality to changes in the same outcome in another locality that did not implement the policy.[3,4] Although the dynamics of COVID-19, individuals' reactions, and the flood of policy responses can hinder the development of credible DiD designs, carefully conducted DiD analyses can be transparent, convincing, timely, and policy relevant. The article by Chae and Park in this issue of *AJPH* titled "Effectiveness of Penalties for Lockdown Violations During the COVID-19 Pandemic in Germany" (p. 1844) is timely and demonstrates how the DiD design can be used to estimate the effects of fines on COVID-19 transmission and mortality rate. Although the study illustrates strengths of the DiD design, it also exposes several pitfalls to applying the DiD design in the context of COVID-19 and NPIs.

SEVEN PITFALLS

Threats to establishing causality and implementing credible DiD designs have been described in earlier issues of this journal.[5] Central to these is establishing a counterfactual (i.e., what the outcomes would have been in the absence of the policy). In the DiD design, this is referred to as the common (parallel) trends assumption, potential violations of which can produce biased causal inference. The dynamics of COVID-19 transmission and the likelihood that policies have time-varying effects are potential threats to the DiD design. Hence, observation time in the pre- and postpolicy periods must be sufficiently long to demonstrate that the parallel trend assumption is robust to confidently conclude that the control and intervention groups were similar before the implementation of the intervention. Decisions about the relevant outcomes and how they are measured (e.g., cases, rates, logarithms) can also affect analyses of COVID-19 policies; for example, the common trends assumption depends on scaling, so what holds at the outcome level does not hold for the log of the outcome.[6]

Second, reverse causality may threaten the credibility of DiD designs. If jurisdictions enact more restrictive policies in response to worsening outbreaks, the observed variations in COVID-19 policies may be functions of past changes in COVID-19 itself; thus small differences at the infection's outset that trigger local interventions may imply large differences in the infection's subsequent development.

Third, voluntary precaution and anticipation may contribute to biased DiD estimates of policy effects. Even before any policy is implemented, publicity on the pandemic may have induced voluntary precautionary behaviors. If worse infections

ABOUT THE AUTHORS

Emmanuel F. Drabo is an assistant professor with and So-Yeon Kang is a PhD student in the Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD. Cynthia L. Gong is a research assistant professor with the Department of Pediatrics, Children's Hospital Los Angeles, Keck School of Medicine of the University of Southern California, Los Angeles.

Correspondence should be sent to Emmanuel Fulgence Drabo, PhD, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, 624 North Broadway St, Suite 402, Baltimore, MD 21205 (e-mail: edrabo@jhu.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

induced stronger precaution and policy interventions, this could exacerbate the bias in the estimated DiD effects. If, in addition, individuals change behaviors in anticipation of the policy implementation, a surge in pre-policy infections may occur, further biasing causal inference.

Fourth, migrations between areas that implemented a given policy and those that did not can contaminate the treatment or control groups and potentially bias the effect size toward null.

Fifth, variable policy timing can lead to staggered treatment and produce time-varying treatment effects, so two-way fixed effects estimates will still be biased away from the estimate of the true treatment effect.[7]

Sixth, inherent measurement errors in COVID-19 outcomes (e.g., reported infections and deaths), which may vary across jurisdictions and over time and lag in time, can further confound estimation of the policy effects. For example, COVID-19's incubation period suggests that reported infections lag true infections by several days, so the epidemiological, social, and economic effects of policies may also lag; differences in timing and use of testing across jurisdictions could exacerbate these biases.

Seventh, the concomitant adoption of multiple COVID-19 prevention strategies makes it particularly difficult to isolate the effects of individual policies; heterogeneity in individual responses to COVID-19 and compliance with policy may produce different effects for various population subgroups and places.

MITIGATING THESE THREATS TO VALIDITY

First, event studies can help detect many of the potential

biases we have described, including ruling out reverse causality if COVID-19 outcomes worsen before the implementation of the policy.

Second, the control group should differ from the treatment group only by the introduction of a single or a few COVID-19 policies. Propensity score reweighting or synthetic control[3] techniques can help balance the treatment and control groups in prepolicy levels and trends in COVID-19 outcomes as well as characteristics that influence disease transmission and severity, individual behaviors, and compliance with policy.

Third, to mitigate the risk that DiD estimates can still be biased when treatment effects are time varying,[7] researchers should consider alternative estimators that rely on the assumption that the potential outcomes are independent of treatment status, conditional on past outcomes (e.g., propensity score reweighting, synthetic control, or lagged dependent variable estimators).

Fourth, although it may be impossible to purge every source of bias, careful consideration of potential sources of biases, their signs, and their likely magnitudes can inform the interpretation of DiD estimates and lend credence to the analysis.

Finally, placebo (falsification) tests should be performed to demonstrate that any observed relationship between the policy and outcomes is more likely attributable to the policy than to other underlying causes. Researchers must recognize these threats to validity to guard against bias, accurately interpret results, and provide sound guidance to policymakers and the public. To avoid overgeneralizations that may lead to unnecessarily harsh measures or further exacerbate existing health and economic

disparities, heterogeneity in policy effects should also be carefully examined.

To tackle the COVID-19 pandemic, various NPI strategies were implemented worldwide, sometimes without carefully weighing the tradeoffs between their health, social, and economic benefits and costs. Now is the time to apply robust, transparent study designs that account for population heterogeneity to credibly assess the impact of these interventions, improve our response to the current pandemic, and, ultimately, prepare for future ones. AJPH

Emmanuel F. Drabo, PhD
So-Yeon Kang, MBA, MPH
Cynthia L. Gong, PharmD, PhD

REFERENCES
1. Flaxman S, Mishra S, Gandy A, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature.* 2020; 584(7820):257–261. https://doi.org/10.1038/s41586-020-2405-7

2. International Monetary Fund. *World Economic Outlook: The Great Lockdown.* Washington, DC; 2020. Available at: https://www.imf.org/en/Publications/WEO/Issues/2020/04/14/weo-april-2020. Accessed September 25, 2020.

3. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J Am Stat Assoc.* 2010;105(490):493–505. https://doi.org/10.1198/jasa.2009.ap08746

4. Angrist JD, Pischke J-S. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press; 2008. https://doi.org/10.2307/j.ctvcm4j72

5. Spiegelman D, Zhou X. Evaluating public health interventions: 8. Causal inference for time-invariant interventions. *Am J Public Health.* 2018;108(9):1187–1190. https://doi.org/10.2105/AJPH.2018.304530

6. Spiegelman D, Khudyakov P, Wang M, Vanderweele TJ. Evaluating public health

interventions: 7. Let the subject matter choose the effect measure: ratio, difference, or something else entirely. *Am J Public Health.* 2018;108(1):73–76. https://doi.org/10.2105/AJPH.2017.304105

7. Goodman-Bacon A. *Difference-in-Differences With Variation in Treatment Timing.* Cambridge, MA: National Bureau of Economic Research; 2018. NBER working paper no. 25018. https://doi.org/10.3386/w25018