

Circles on the grounds of San Francisco's Dolores Park are designed to limit the spread of SARS-CoV-2 by encouraging social distancing.

SOCIAL SCIENCE: COVID-19

Which interventions work best in a pandemic?

We can exploit randomized controlled trials, compartmental models, and spillovers

By Johannes Haushofer^{1,2,3,4,5} and C. Jessica E. Metcalf⁶

he only approaches currently available to reduce transmission of the novel coronavirus severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) are behavioral: handwashing, cough and sneeze etiquette, and, above all, social distancing. Policymakers have a variety of tools to enable these "nonpharmaceutical interventions" (NPIs), ranging from simple encouragement and recommendations to full-on regulation and sanctions. However, these interventions are often used without rigorous empirical evidence: They make sense in theory, and mathematical models can be used to predict their likely impact (1, 2), but with different policies being tried in different places—often in complicated combinations and without systematic, built-in evaluation—we cannot confidently attribute any given reduction in transmission to a specific policy.

Because many of these interventions differ from each other in terms of their economic and psychological cost-ranging from very inexpensive, in the case of interventions based on behavioral economics and psychology, to extremely costly, in the case of school and business closures-it is crucial to identify the interventions that most reduce transmission at the lowest economic and psychological cost. Randomized controlled trials (RCTs) are one of several methods that can be used for this purpose but surprisingly have received little attention in the current pandemic, despite a long history in epidemiology and social science. We describe how RCTs for NPIs can be practically and ethically implemented in a pandemic, how compartmental models from infectious disease epidemiology can be used to minimize measurement requirements, and how to control for spillover effects and harness their benefits.

JUSTIFIABLE RCTS

How can RCTs be practically and ethically conducted in a pandemic? In a typical RCT. a subset of randomly chosen individuals or regions receives an intervention, and a randomly chosen control group receives no intervention or a different intervention. The random assignment ensures that any later differences between the groups can be attributed to the intervention. During an outbreak, policy-makers must decide which interventions to impose when, and when to loosen them again. It will rarely be feasible in this context to omit individuals or regions entirely. However, policy-makers can use systematic timing of such interventions to both protect the population and understand the impact of the intervention. For example, when experts begin to think that measures can be loosened, this can be done gradually, so that evaluation is possible: A subset of randomly chosen locations (such as counties or municipalities) begins, and others gradually follow suit. Comparison of the "early" to the "late" regions makes it possible to estimate the effects of the intervention.

This "phase-in" or "stepped-wedge" approach can be used at any point during the

¹Department of Psychology and Woodrow Wilson School of Public Affairs, Princeton University, Princeton, NJ, USA.³National Bureau of Economic Research, Cambridge, MA, USA.³Busara Center for Behavioral Economics, Nairobi, Kenya.⁴Center for Global Development, Washington, DC, USA.⁵Max Planck Institute for Collective Goods, Bonn, Germany.⁵Department of Ecology and Evolutionary Biology and Woodrow Wilson School of Public Affairs, Princeton University, Princeton, NJ, USA. Email: cmetcalf@princeton. edu; haushofer@princeton.edu

pandemic. At the beginning, protective measures can begin early in some areas and somewhat later in others. During the pandemic, periods of loosened measures may be necessary to restore a sense of normality and keep essential services working, or measures may have to be tightened to limit further spread of the virus; these periods can also be systematically timed to evaluate their impact. In extended versions, different interventions can be tested against each other, and different locations can tighten or loosen different subsets of restrictions; for example, schools could be opened back up. whereas nonessential businesses remain closed.

Governments and organizations could work with scientists to choose an experimental design, implement and keep track of the treatment assignment. and measure outcomes. Studies of this kind can now often be done in nimble and practicable ways, reducing the oversight and time burden on implementpartners. Interventions ing could range from messaging campaigns to promote social distancing to laws and regulations. Where full randomization (without phase-in) is possible, this may be desirable to increase statistical power (3).

RCTs are, of course, not the only method for estimating the impact of NPIs. Where ran-

domization is not feasible, the "natural experiments" created by some policies can be exploited, such as quasi-arbitrary cutoffs (for example, the reopening of stores below a certain square footage). Observational studies, often integrated with mathematical models have also contributed important insights.

Great care must be exercised to make RCTs ethical. Several considerations are relevant: The approach may be ethically justifiable because there are two sources of uncertainty around most interventions. For any intervention, it may be uncertain whether its benefits in terms of reducing disease transmission exceed its economic and psychological costs or how these costs and benefits relate to those of other interventions. At the same time, it is difficult to identify a single "correct" moment to loosen or tighten protective measures, as illustrated by ongoing policy debates. Thus,

Testing interventions during an outbreak

Time course of infection in the absence of an intervention (red) and with an intervention (blue) that is either loosened (top) or tightened (middle) for 2 weeks. The number of cases at the end of different tightening or loosening windows (bounds indicated with dashed vertical lines) in regions where the intervention was tightened or loosened compared with regions where it remained in place provides a measure of the relative change in transmission associated with the intervention. Final size (bottom) is affected by the time point at which the 2-week tightening or loosening of the intervention is initiated (5). See supplementary materials for details.



equipoise may be satisfied in terms of costs, benefits, and timing. Policy-makers are therefore neither knowingly withholding a beneficial intervention from constituents nor knowingly imposing a harmful one. This uncertainty is likely to make staggered tightening or loosening of an intervention more acceptable to the public.

Further, the phase-in or stepped-wedge approach may be ethically justifiable because individuals in both control and treatment groups eventually experience the costs and benefits of any intervention. In addition, even short periods of tightening or loosening can be used to determine the impact of mitigation measures, minimizing the burden on whichever group experiences the smaller benefits. A powerful illustration of the ethical acceptability of this phase-in approach among both scientists and the public is its use in RCTs of vaccines, even for highly lethal pathogens such as Ebola (4).

MODELS TO GUIDE DATA COLLECTION

Careful measurement of outcomes is crucial for this approach to succeed. In particular, it is essential to understand the impact of any given intervention on the full epidemic trajectory [see supplementary materials (SM)]. However, the requirements measurement can be simplified if data collection and analysis are guided by compartmental models from infectious disease epidemiology. The time course of infections is affected in a SIR model (reflecting the three possible states of an individual in the community: susceptible, infectious, or recovered) when one group of locations (such as counties or districts) loosens or tightens the intervention for 2 weeks while another maintains the status quo (see the figure) (5). Crucially, because the SIR model describes the entire trajectory of an outbreak using only two parameters, very few measurements are required to estimate them. In particular, using only estimates of the number of infections at the end of the intervention in treatment and control regions, we can estimate how much a given intervention reduces transmission relative to no intervention (see SM). In addition, this difference allows policy-makers to determine which of several interven-

tions reduced transmission the most and by how much. If additional information about the number of infections at the beginning of the intervention is available, we can further estimate whether transmission has been sufficiently reduced that the outbreak is shrinking (corresponding to an effective reproductive number below 1).

Insights from epidemiology can also be used to address several additional questions: In addition to learning how much an intervention changes the transmission rate, policy-makers may also want to know how different interventions affect the "final size" of the pandemic-what share of the population will have been infected in total when the pandemic has died down. Also, they may want to understand how a single intervention might perform if it were deployed at different time points during the pandemic (for example, early versus late) but can only test it once. Additionally, they may

GRAPHIC: X. LIU/SCIENCE

want to directly compare the effectiveness of two interventions, despite them having been deployed at different times, because it may not always be possible to time periods of tightening and loosening precisely.

In the stylized model, all of these estimates can be derived from adding a single measurement at one time point to those described above—namely, the number of susceptibles [measured, for example, with serology (6)] (see SM). Of course, the available capacity for polymerase chain reaction and serology needs to be able to support such studies, but testing capacity is growing around the world, suggesting feasibility.

An important caveat to reducing the measurement requirements is that the above approach leans on the assumptions of a relatively simple SIR model; in particular, both the transmission rate and the impact of an intervention on this rate are assumed to remain constant throughout the pandemic. However, it is straightforward to extend the model to accommodate inherent variation in transmission through time, and complex treatment effects that may be a signature of NPIs, including decay over time (for example, fatigue from a lockdown or fading response to a messaging campaign), persistence (for example, hygiene behaviors such as handwashing which turn into a habit), or intensification over time (such as messaging campaigns that "go viral"). In such cases, the measurement requirements will increase to identify the additional parameters (such as decay) contained in the extended model. Similarly, the basic compartmental model can be extended [to SIRS (susceptible-infected-recovered-susceptible), SEIR (susceptible-exposed-infectious-recovered), or age-structured models, for example] to reflect additional features of the transmission process (duration of immunity, latent period, or variable contact patterns over age) or the intervention (for example, if it targets specific age groups).

Thus, the effects of interventions on disease transmission can be estimated with the help of epidemiological models. However, the economic and psychological costs and benefits of such interventions are equally important. Reducing the number of measurements by leveraging the SIR model is not possible for these outcomes, about which the model makes no predictions and whose time course need not follow that of infections. For example, a "successful" intervention that reduces the risk of overburdening the health system will have the effect of spreading the infections over time. This implies that the desirable behaviors induced by any intervention have to be maintained for longer to outlast the duration of the pandemic. This may impose psychological and economic costs on the population that are larger than those that would be incurred in a more temporally condensed pandemic. In the absence of a model, these effects can only be captured with careful measurement over time.

SPILLOVER EFFECTS

Interventions delivered to some regions or individuals but not others are likely to nevertheless affect those who were not targeted. Such so-called "spillover" effects present both a challenge and an opportunity in evaluating the impact of NPIs. The opportunity is that such spillovers can generate strongly increasing returns to intervention coverage in terms of individual protection (7): they can therefore be harnessed to maximize the effects of a given intervention. For example, consider a hypothetical intervention that reduces the size of a pandemic by 15% when it is delivered to 20% of a community. Because of the nonlinear dynamics of infection that arise from depletion of the number of those susceptible to infection, increasing the coverage to 60% may generate a greater-than-proportional reduction in pandemic size of 56%.

At the same time, such spillovers pose a challenge to the estimation of treatment effects. However, standard trial designs are available to enable measurement of spillovers (8-10). In particular, nonlinear returns to saturation (the share of the population exposed to an intervention) can be integrated into tests of interventions by creating variation in spatial saturation of intervention delivery. For example, groups of 15 locations might be randomized to a "low saturation" condition in which a third of locations are treated with an intervention-for example, the distribution of face masks or hand sanitizer, or opening or closing of parks or schools-or to a "high saturation" condition, in which two-thirds of locations are treated. Such studies have to be relatively large scale to achieve adequate statistical power; power calculations are therefore important, and using more than two or three levels of saturation may not be practicable.

Because spatial spillovers may occur at different spatial scales, causal inference methods that flexibly allow for such complications have to be used. Data on the source of spillovers, such as the commuting patterns of essential workers, can help identify relevant spatial scales. The feasibility of this approach in terms of both statistical power and causal inference in the presence of spillovers of unknown spatial dimensions has been suggested by recent large-scale studies on the general equilibrium effects of economic interventions (*11*). Thus, tests of interventions to combat COVID-19 should take advantage of, and measure, these nonlinear effects of saturation.

NPIs can be rigorously tested by using randomization without compromising scientific and ethical standards. Although this approach will require more time than generating projections from observational methods and mathematical models, the benefits in terms of accuracy could be considerable. If policy-makers and scientists combine insights from infectious disease epidemiology with carefully and ethically designed impact evaluation, alongside other empirical and theoretical methods for studying impact (12–14), they will have a powerful tool for reducing the human health, societal, and economic costs in the SARS-CoV-2 pandemic and in pandemics in general.

REFERENCES AND NOTES

- N. Ferguson et al., "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand," Working paper, Imperial College COVID-19 Response Team, London, UK, 16 March 2020.
- T. Rawson et al., medRxiv 20084517 [Preprint] 2 May 2020; https://doi.org/10.1101/2020.04.29.20084517
- L. Kennedy-Shaffer, M. Lipsitch, medRXiv 20087429 [Preprint] 6 May 2020; https://doi.org/10.1101/2020.0 5.01.20087429
- A. M. Henao-Restrepo et al., Lancet 389, 505 (2017). doi:10.1016/S0140-6736(16)32621-6 Medline
- F. Di Lauroy, I. Z. Kissy, J. C. Miller, "The timing of oneshot interventions for epidemic control," Working paper, La Trobe University, Melboourne, Australia.
- C. J. Metcalf *et al.*, *Lancet* 388, 728 (2016). doi:10.1016/ S0140-6736(16)30164-7 Medline
- R. M. Anderson, R. M. May, Infectious Diseases of Humans: Dynamics and Control (Oxford Univ. Press, 1992).
- 8. J. Benjamin-Chung *et al.*, *Int. J. Epidemiol.* **47**, 332 (2018); doi:10.1093/ije/dyx201 Medline
- M. E. Halloran, C. J. Struchiner, *Epidemiology* 2, 331 (1991); doi:10.1097/00001648-199109000-00004 Medline
- 10. E. Miguel, M. Kremer, Econometrica **72**, 159 (2004). doi:10.1111/j.1468-0262.2004.00481.x
- D. Egger *et al.*, "General equilibrium effects of unconditional cash transfers: Experimental evidence from Kenya," Working paper26600, NBER, Cambridge, MA, 21 November 2019.
- H. Fang, L. Wang, Y. Yang, "Human mobility restrictions and the spread of the novel coronavirus (2019-nCoV) in China," Working paper 26906, National Bureau of Economic Research (NBER), Cambridge, MA, March 2020.
- S. Flaxman et al., "Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries, Working paper, Imperial College COVID-19 Response Team, London, UK, 30 March 2020.
- H. Tian et al., Science 368, 638 (2020); doi:10.1126/ science.abb6105 Medline

ACKNOWLEDGMENTS

We thank D. Björkegren, A. Chandrasekhar, C. de Chaisemartin, J. de Quidt, B. Grenfell, R. Hussam, S. Jayachandran, D. Strömberg, and anonymous referees for helpful comments and suggestions.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1063/suppl/DC1

Published online 21 May 2020 10.1126/science.abb6144



Which interventions work best in a pandemic?

Johannes Haushofer and C. Jessica E. Metcalf

Science **368** (6495), 1063-1065. DOI: 10.1126/science.abb6144originally published online May 21, 2020

ARTICLE TOOLS	http://science.sciencemag.org/content/368/6495/1063
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2020/05/20/science.abb6144.DC1
RELATED CONTENT	http://stm.sciencemag.org/content/scitransmed/9/396/eaal3653.full http://stm.sciencemag.org/content/scitransmed/12/541/eabb5883.full http://stm.sciencemag.org/content/scitransmed/11/499/eaat0360.full http://stm.sciencemag.org/content/scitransmed/12/534/eabb1469.full http://stm.sciencemag.org/content/scitransmed/12/549/eabb9401.full
REFERENCES	This article cites 6 articles, 1 of which you can access for free http://science.sciencemag.org/content/368/6495/1063#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2020, American Association for the Advancement of Science