



Measuring the scientific effectiveness of contact tracing: Evidence from a natural experiment

Thiemo Fetzer^a and Thomas Graeber^{b,1}

^aCompetitive Advantage in the Global Economy Research Centre, Department of Economics, University of Warwick, Coventry CV4 7AL, United Kingdom; and ^bHarvard Business School, Harvard University, Boston, MA 02163

Edited by Alan Hastings, University of California, Davis, CA, and approved May 25, 2021 (received for review January 14, 2021)

Contact tracing has for decades been a cornerstone of the public health approach to epidemics, including Ebola, severe acute respiratory syndrome, and now COVID-19. It has not yet been possible, however, to causally assess the method's effectiveness using a randomized controlled trial of the sort familiar throughout other areas of science. This study provides evidence that comes close to that ideal. It exploits a large-scale natural experiment that occurred by accident in England in late September 2020. Because of a coding error involving spreadsheet data used by the health authorities, a total of 15,841 COVID-19 cases (around 20% of all cases) failed to have timely contact tracing. By chance, some areas of England were much more severely affected than others. This study finds that the random breakdown of contact tracing led to more illness and death. Conservative causal estimates imply that, relative to cases that were initially missed by the contact tracing system, cases subject to proper contact tracing were associated with a reduction in subsequent new infections of 63% and a reduction in subsequent COVID-19-related deaths of 66% across the 6 wk following the data glitch.

COVID-19 | contact tracing | public health | infectious diseases

Contact tracing has been a central pillar of the public health response to COVID-19, with countries around the world allocating unprecedented levels of resources to the build-up of their testing and tracing capacities (1). Public health experts argue that even as vaccines have become available, nonpharmaceutical interventions such as contact tracing remain indispensable (2). Simultaneously, however, the effectiveness of contact tracing has been subject to controversial public and scientific debates: reports on low adherence to self-quarantine, insufficiently trained contact tracers, and people providing incomplete or inaccurate information about their contacts due to concerns about privacy, stigma, and scams abound (3–8). Why does significant uncertainty about the effectiveness of contact tracing persist?

One reason is that the type of evidence required for its evaluation is notoriously hard to obtain. Ideally, public policies are based on causal evidence demonstrating their effectiveness, which requires randomized experiments. Experimenting with public policies, however, is often infeasible due to logistical constraints and ethical concerns. For example, it may not be morally acceptable to implement better contact tracing in some randomly selected areas than in others, because this may selectively lead to more adverse outcomes in specific areas. As a consequence, scientific research on the effectiveness of contact tracing both in previous pandemics (10–13) and during COVID-19 (9, 14–19) has had to rely on observational data and modeling techniques. The existing correlational evidence points to a positive impact of contact tracing measures but is subject to the concern that correlations may not reflect a causal relationship. For example, the underlying variation in contact tracing may co-occur with changes in other policies such as contact restrictions or with changing epidemiological trends, so that it is difficult to cleanly identify which factor is truly responsible for an observed correlation.

To address this lack of causal evidence, we exploit a unique source of experimental variation in contact tracing. On October 4, 2020, the public health authorities in England released a public statement

on a “technical issue” discovered in the night of October 2 to October 3 (20). An internal investigation had revealed that a total of 15,841 positive cases had accidentally been missed in both the officially reported figures and the case data that was transferred to the national contact tracing system—around 20% of all cases during that time. This omission occurred because case information had accidentally been truncated from Excel spreadsheets after a row limit had been reached. According to government reports, the original reporting dates of the missed cases would have been between September 25 and October 2. While the data glitch did not affect the individual dissemination of test results to people who tested positive, an anticipated 48,000 close recent contacts had not been traced in a timely manner and had therefore not been ordered to self-quarantine. The evolution of the daily number of newly reported cases in England is shown in Fig. 1A (black line). The reporting date of a case simultaneously marks the day on which it is referred to the contact tracing system (21, 22). The figure further shows the number of positive test results by the date on which these tests were actually taken (i.e., their so-called specimen date [red line]). Reported cases (black line) trail behind actual cases (red line) due to a natural lag in reporting: because laboratory tests need to be evaluated and processed, close to 100% of all new test results enter the official statistics and are referred to the contact tracing with a delay of two to five days (see also *SI Appendix*, Fig. S1 and Table S1). Fig. 1A documents a striking change in the relationship between reported and actual cases during the time of the data glitch. Reported cases moved

Significance

Contact tracing constitutes the backbone of nonpharmaceutical public interventions against COVID-19, as it did with previous pandemics. Experts argue that its importance rises again as vaccination rates increase and the spread of COVID-19 slows, which makes tracing of individual cases possible. However, because randomized experiments on contact tracing are infeasible, causal evidence about its effectiveness is missing. This shortage of evidence is alarming as governments around the world invest in large-scale contact tracing systems, frequently facing a lack of cooperation from the population. Exploiting a large-scale natural experiment, we provide evidence that contact tracing may be even more effective than indicated by previous correlational research. Our findings inform current and future public health responses to the spread of infectious diseases.

Author contributions: T.F. and T.G. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

See [online](#) for related content such as Commentaries.

¹To whom correspondence may be addressed. Email: tgraeber@hbs.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2100814118/-DCSupplemental>.

Published August 12, 2021.

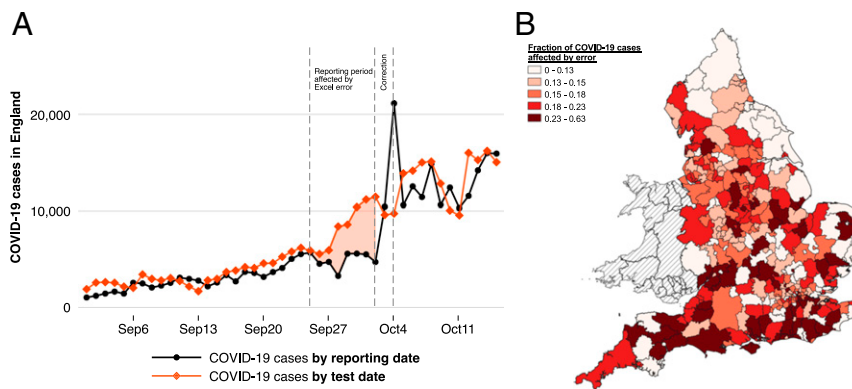


Fig. 1. Evolution of COVID-19 in England and regional variation in contact tracing delays due to the Excel error. (A) COVID-19 cases in England separately by date of test and by reporting date. The reporting date equals the date on which a case is referred to the national contact tracing system. Reported cases trail behind actual positive case numbers due to a normal reporting lag. Reported and actual cases notably diverge during the time in which the Excel error occurred, highlighted by the area shaded in red. (B) For each local authority district, we calculate the fraction of all local COVID-19 cases with test dates between September 20 and September 27, 2020, that were referred to contact tracing with an unusual delay of 6 to 14 d due to the Excel error. The different color shades represent different quintiles of the distribution of this fraction measure. The map shows substantive heterogeneity in how strongly different areas were affected.

laterally, whereas true cases surged dramatically, leading to a much more pronounced divergence. From just 4,726 cases on October 2, the number of reported cases skyrocketed to 10,436 on October 3 and 21,140 on October 4, which is when the correction occurred. This development was accompanied by a notable worsening in the performance of contact tracing; see *SI Appendix, Fig. S2*. While the share of contacts reached within the first 24 h hovered steadily around 80% in prior weeks, it plummeted to just above 60% when the delays in contact tracing kicked in.

The Excel error created a natural experiment that allows us to study the causal effect of contact tracing: by chance, the loss of case information was much more severe for some areas of England than others. This source of random variation allows us to investigate to which extent areas that were more strongly affected by the lack of contact tracing subsequently experienced a different pandemic progression.

We are able to reconstruct the local number of delays in contact that are due to the Excel error using a simple approach that relies on linkages between historically published daily data sets on newly reported cases. In the period leading up to the data glitch, an average of 96% of all positive test results with a given specimen date had been reported and referred to contact tracing within the five days following the test date (*SI Appendix, Table S1*). This information on normal reporting lags allows us to identify, at the local level, the number of positive test results that were reported and referred to contact tracing unusually late. Specifically, our simplest measure of missed cases exploits only test dates that precede the discovery of the Excel error (October 3) by so long that, under normal circumstances, virtually all positive cases should have been reported before that correction date. For each of the 315 Lower Tier Local Authorities (LTLA) in England, we sum up the number of positive results from tests taken between September 20 and September 27 that were reported on or after October 3. This number, expressed as a share of all local cases with the same specimen dates, yields a percentage indicator of how strongly the data glitch affected each LTLA through delays in contact tracing. The map displayed in Fig. 1B documents substantial variation in the geographic signature of the Excel error. This variation is unrelated to area-specific characteristics, such as its population size or its previous exposure to and the recent trend in COVID-19 spread. An analysis of covariate balance (*SI Appendix, Table S14*) reveals no systematic relationships between measures of missed cases and area characteristics. Note that our baseline measure is conservative to the extent that it only relies on

early specimen dates for which delays due to the Excel error can be inferred with high confidence, and it thus underestimates the officially reported extent of the breakdown in contact tracing. All of our findings are robust to alternative approaches of calculating the local number of delays in contact tracing (*SI Appendix*).

Results

Before quantifying the causal effect of delays in contact tracing, we illustrate the relationship between delays in contact tracing and COVID-19 infection dynamics. Note that the local share of cases subject to contact tracing delays provides a continuous measure of local affectedness by the Excel error. For the purpose of a visual illustration, we now split all 315 LTLAs into two groups based on whether contact tracing in an area was relatively strongly affected or relatively little affected. We separate areas based on whether their share of cases subject to delays in contact tracing was above or below the median share. Fig. 2 plots the average of COVID-19 incidence per 100,000 population separately for areas with above-median shares of cases subject to delays in contact tracing and areas with below-median shares. The data are plotted for specimen dates at the weekly level. We make three observations. First, the two groups experienced virtually identical epidemiological trajectories in the weeks preceding the onset of the data glitch. Second, we see an increase in COVID-19 incidence across time in both groups. Third, and most importantly, the increase in COVID-19 infections was much more pronounced in areas with above-median exposure to delays in contact tracing. This divergence started during the period of the data glitch and led to a quantitatively large difference in infection intensity across the four weeks following the data error. Except for the divergence appearing during the time of the data glitch, the development of COVID-19 incidence over time looked remarkably similar between the two groups. This again reflects the random nature of the variation in how strongly different areas were affected.

To provide a quantitative estimate of the causal effect of delays in contact tracing, we follow a canonical “difference-in-differences” regression approach. Crucially, this empirical strategy is immune to the fact that the COVID-19 development across England already displayed an upward trend before the error occurred: the effect estimate subtracts out the “normal” trend in COVID-19 spread in areas that were not affected by the Excel error. Intuitively, for each area, the estimation first computes the difference in the spread of COVID-19 before and after the data glitch. The

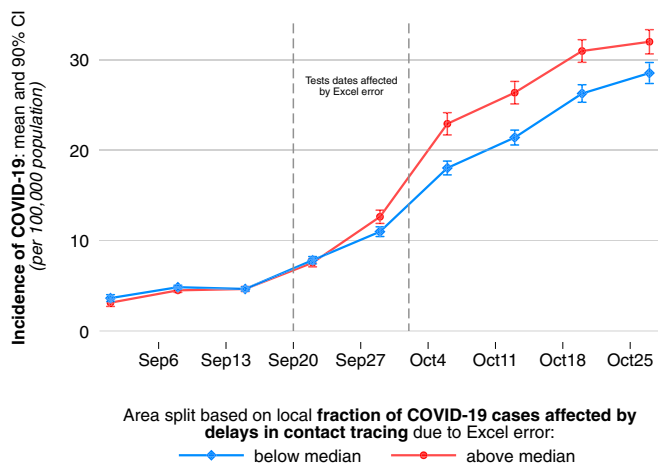


Fig. 2. Evolution of local COVID-19 incidence in areas with above median versus areas with below median exposure to delays in contact tracing due to the Excel error. For each of the 315 LTLA in England, we calculate the share of positive COVID-19 tests taken between September 20 and September 27 that were referred to contact tracing with an unusual delay of 6 to 14 d due to the Excel error. We create two equally sized groups of areas based on whether they—by chance—experienced above median or below median exposure to unusual delays in contract tracing. We plot the average incidence of COVID-19 for each group by test date. We observe virtually identical pretreatment trends across groups but a substantive and persistent divergence in COVID-19 spread at the onset of the period during which the Excel error occurred, which is highlighted by the dashed lines; 90% CIs are displayed.

difference-in-differences estimator captures to what extent this local change over time differs between areas that were, by chance, more strongly affected by delays in contact tracing and areas that were less strongly affected. Note that this approach relies on cross-area comparisons and is thus immune to nationwide epidemiological trends in infections that affect all regions similarly. Our local measure of exposure to delays in contact tracing due to the Excel error is M_i . M_i captures the number of late referrals per 100,000 population in area i that was likely due to the data glitch (see *SI Appendix* for further details on how M_i is constructed). The resulting baseline regression specification is

$$y_{i,t} = \mu_i + \gamma_t + \eta \times \text{Post}_t \times M_i + \beta' X_{i,t} + \epsilon,$$

where $y_{i,t}$ denotes a measure of COVID-19 spread in area i on day t . We study various outcome measures $y_{i,t}$. The regression controls for area fixed effects, μ_i , and day fixed effects, γ_t . Across

different specifications, we also control for a host of additional measures, $X_{i,t}$, that account for, first, the nonlinear nature of case growth by controlling for previous levels and trends in COVID-19 spread and, second, for a multiplicity of more than 50 area characteristics (see also the covariate balance *SI Appendix*, Table S14). The area characteristics include employment shares in one-digit industries, educational attainment, socio-economic status of the resident population, which also captures shares in full time education or in university, and regular in- and out-commuting flows. These time-invariant measures are interacted with a set of date fixed effects to account for potential nonlinear growth. However, we do not expect those controls to significantly affect our estimate of the coefficient of interest, η , due to the random nature of exposure to delays to contact tracing M_i . In Fig. 3, we show estimation results for η , capturing the average effect of delays to contact tracing on new infections and new COVID-19 deaths based on the above specification. The full results are reported in *SI Appendix*, Table S2. Our conservative estimate for the effect of delays in contact tracing on daily new infections during calendar weeks 39 through 44 is 0.443. This implies that one additional case referred late to contact tracing was associated with a total of 18.6 additional infections reported in the 6-wk posttreatment period under consideration. Similarly, our estimation yields a precisely estimated and significant effect of 0.006 additional daily deaths per late referral. This corresponds to around 0.24 new COVID-19 deaths per late referral to contact tracing overall in the following 6 wk.

To put these effect sizes into perspective, we emphasize that our estimates capture the cumulative effect of late referrals over a 6-wk period. Given a total of 43,875 positive cases with test dates during September 20 through 27 and a total of 597,381 cases in the 6-wk posttreatment period, the raw case data imply that, on average, each case between September 20 and 27 was followed by an expected 13.6 new cases across the following 6 wk. Our analysis allows us to disentangle between this statistical multiplier for a case referred to contact tracing with an Excel error-induced delay to the multiplier for a case that was traced as normal. Our regression estimate implies that proper, timely contact tracing during September 20 through 27 is associated with a 63% smaller multiplier (from 29.4 to 10.8) relative to contact tracing that was subject to unusual delays of between 6 and 14 d. The same type of regression estimate for COVID-19-related deaths implies that timely contact tracing was associated with 66% fewer deaths (compared to contact tracing with unusual delays of between 6 and 14 d) across the 6-wk period following September 20 through 27.

SI Appendix, Figs. S5 and S6 and Tables S3 and S4 shed light on the potential epidemiological mechanisms associated with these findings, showing that the increase in infections and deaths

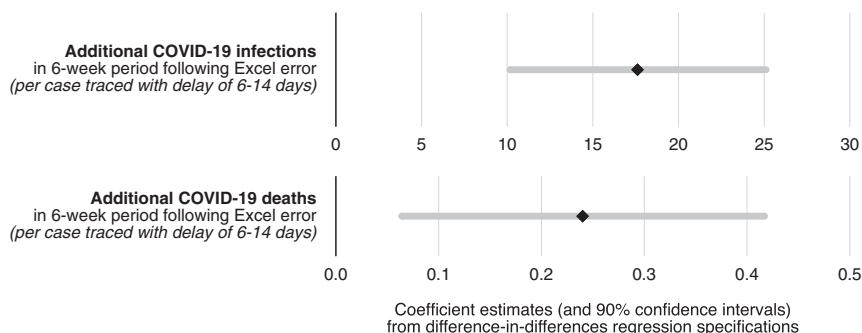


Fig. 3. Impact of delays in contact tracing on new COVID-19 infections and deaths. Difference-in-differences regression estimates (at level of LTLA) for the average effect of each case referred to contact tracing with a delay of between 6 and 14 d on new infections per capita and new COVID-19 deaths per capita in the 6 wk period following the Excel error. All regressions control for area fixed effects and date fixed effects, along with nonlinear local time trends in pretreatment infection intensity as well as nonlinear trends in a host of over 50 area characteristics. Complete regression results are reported in *SI Appendix*, Table S2; 90% CIs are displayed.

was accompanied by an increase in the test positivity rate, a sharp increase in number of tests performed and a worsening of the quality of contact tracing. We further find that the effects are robust i) to different ways of constructing the measure of delays in contact tracing, ii) to the level of spatial disaggregation, indicating no significant role of interregional spillover effects, iii) to the exclusion of individual regions, iv) to alternative empirical strategies such as one based on matching areas that had been evolving highly similarly in terms of the pandemic development prior to the data glitch, v) to alternative functional forms of the estimated relationship (e.g., log-log specifications) that account the nonlinear nature of infection dynamics differently, vi) to alternative ways of conducting statistical inference, specifically, randomization inference, and vii) we conduct empirically highly conservative placebo tests (*SI Appendix*).

Across this set of analyses, our point estimates of the effect imply that the specific failure of timely contact tracing due to the Excel error is associated with between 126,836 (22.5% of all cases in the 6-wk period following the discovery of the error) and 185,188 (32.8%) additional reported infections, and with between 1,521 (30.6% of all deaths) and 2,049 (41.2%) additional COVID-19-related reported deaths (*SI Appendix*).

Discussion

Contact tracing has repeatedly attracted criticism that partly reflects a shortage of causal evidence for its effectiveness. Reliable evaluations of public health interventions are in dire need because novel policy measures can have unintended harmful consequences (23). This study delivers a casual analysis, showcasing how empirical research can help evaluate public health policies by exploiting natural experiments. The findings complement the state of existing correlational evidence: despite the multiplicity of challenges that contact tracing faces in practice, this nonpharmaceutical intervention can have a strong impact on the progression of a pandemic. The estimated effect sizes are notable in light of the baseline delays

that test and trace programs face even in the absence of unusual errors, such as the delay between the onset of illness and the testing date or the time lag between specimen and reporting dates due to test processing times. In the context under consideration, the nontimely referral to contact tracing has likely contributed to propelling England to a different stage of COVID-19 spread at the onset of a second wave. Our findings should be viewed in the specific context of England with a nationally centralized tracing system. Due to the heterogeneity in how populations cooperate with official contact tracing efforts, for example, we do not claim generality of our estimated effect sizes for other countries. The robust and quantitatively large effects estimated under conservative assumptions across our analyses suggest, however, that contact tracing may be an even more effective tool to fight infectious diseases than was previously thought.

Materials and Methods

Our baseline analyses leverage three sources of publicly available data: the United Kingdom's COVID-19 dashboard, statistics on the Test and Trace program published by the NHS, and weekly data on deaths published by the Office for National Statistics. We identify late referrals to contact tracing that are likely due to the data glitch by exploiting the evolution of reported cases numbers for a given specimen date across different reporting dates, as well as a variety of other approaches detailed in the *SI Appendix*.

Data Availability. Previously published data were used for this work (various sources of publicly available COVID-19 data as specified in the text). All data and code are available in GitHub at <https://bit.ly/3rFjnkp>.

ACKNOWLEDGMENTS. We thank Arun Advani, Tim Besley, Mirko Draca, Stephen Hansen, Leander Heldring, Neil Lawrence, Andrew Oswald, Chris Roth, Flavio Toxvaerd, and Yee Whye Teh for helpful comments. Preliminary findings of this study on the effectiveness of contact tracing were made available as preprints (24, 25). T.F. acknowledges support from the Economic and Social Research Council that indirectly supported this research through his affiliation with the Competitive Advantage in the Global Economy Research Centre.

1. A. Desvars-Larrive *et al.*, A structured open dataset of government interventions in response to COVID-19. *Sci. Data* **7**, 285 (2020).
2. N. Ferguson *et al.*, Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and health-care demand. *Imperial College London*. **10**, 77482 (2020).
3. J. Li, X. Guo, COVID-19 contact-tracing apps: A survey on the global deployment and challenges. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2005.03599> (Accessed 1 November 2020).
4. E. Clark, E. Y. Chiao, E. S. Amirian, Why contact tracing efforts have failed to curb coronavirus disease 2019 (COVID-19) transmission in much of the U.S. *Clin. Infect. Dis.* **72**, e415–e419 (2021).
5. H. Cho, D. Ippolito, Y. W. Yu, Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2003.11511> (Accessed 1 November 2020).
6. S. Altmann *et al.*, Acceptability of app-based contact tracing for COVID-19: Cross-country survey study. *JMIR Mhealth Uhealth* **8**, e19857 (2020).
7. R. K. Webster *et al.*, How to improve adherence with quarantine: Rapid review of the evidence. *Public Health* **182**, 163–169 (2020).
8. G. J. Rubin, L. E. Smith, G. J. Melendez-Torres, L. Yardley, Improving adherence to 'test, trace and isolate'. *J. R. Soc. Med.* **113**, 335–338 (2020).
9. Scientific Advisory Group for Emergencies, Summary of the effectiveness and harms of different non-pharmaceutical interventions. <https://www.gov.uk/government/publications/summary-of-the-effectiveness-and-harms-of-different-non-pharmaceutical-interventions-16-september-2020>. Accessed 1 November 2020.
10. C. Fraser, S. Riley, R. M. Anderson, N. M. Ferguson, Factors that make an infectious disease outbreak controllable. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6146–6151 (2004).
11. Y. H. Chan, H. Nishiura, Estimating the protective effect of case isolation with transmission tree reconstruction during the Ebola outbreak in Nigeria, 2014. *J. R. Soc. Interface* **17**, 20200498 (2020).
12. A. Pandey *et al.*, Strategies for containing Ebola in West Africa. *Science* **346**, 991–995 (2014).
13. D. Klinkenberg, C. Fraser, H. Heesterbeek, The effectiveness of contact tracing in emerging epidemics. *PLoS One* **1**, e12 (2006).
14. I. Braithwaite, T. Callender, M. Bullock, R. W. Aldridge, Automated and partly automated contact tracing: A systematic review to inform the control of COVID-19. *Lancet Digit. Health* **2**, e607–e621 (2020).
15. M. Kendall *et al.*, Epidemiological changes on the Isle of Wight after the launch of the NHS Test and Trace programme: A preliminary analysis. *Lancet Digit. Health* **2**, e658–e666 (2020).
16. J. Hellewell *et al.*, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**, e488–e496 (2020).
17. M. E. Kretzschmar *et al.*, Impact of delays on effectiveness of contact tracing strategies for COVID-19: A modelling study. *Lancet Public Health* **5**, e452–e459 (2020).
18. A. J. Kucharski *et al.*, CMMID COVID-19 Working Group, Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: A mathematical modelling study. *Lancet Infect. Dis.* **20**, 1151–1160 (2020).
19. N. Lu, K. W. Cheng, N. Qamar, K. C. Huang, J. A. Johnson, Weathering COVID-19 storm: Successful control measures of five Asian countries. *Am. J. Infect. Control* **48**, 851–852 (2020).
20. Public Health England, PHE statement on delayed reporting of COVID-19 cases. <https://www.gov.uk/government/news/phe-statement-on-delayed-reporting-of-covid-19-cases>. Accessed 1 November 2020.
21. Department of Health & Social Care, COVID-19 testing data: Methodology note. <https://www.gov.uk/government/publications/coronavirus-covid-19-testing-data-methodology/covid-19-testing-data-methodology-note>. Accessed 1 November 2020.
22. Department of Health & Social Care, NHS Test and Trace statistics (England): Methodology. <https://www.gov.uk/government/publications/nhs-test-and-trace-statistics-england-methodology/nhs-test-and-trace-statistics-england-methodology>. Accessed 1 November 2020.
23. T. Fetzter, Subsidizing the spread of COVID-19: Evidence from the UK's Eat-Out-to-Help-Out scheme. <https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/wp.517.2020.pdf>. Accessed 29 July 2021.
24. T. Fetzter, T. Graeber, Learning about the effectiveness of contact tracing from when it failed—A natural experiment from England. <https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/bn30.2020.pdf>. Accessed 29 July 2021.
25. T. Fetzter, T. Graeber, Does contact tracing work? Quasi-experimental evidence from an Excel error in England. *medRxiv [Preprint]* (2020). <https://www.medrxiv.org/content/10.1101/2020.12.10.20247080v1>. Accessed 29 July 2020.