

# Authors Response to Letters to the editor regarding: ‘Assessing mandatory stay- At- Home and business closure effects on the spread of COVID- 19’

We are pleased to see the active discussion around our study on the relationship between mandatory stay-at-home and business closures and COVID-19 spread.<sup>1</sup> In this response, we address issues raised in three letters.<sup>2-4</sup>

## 1 | SAMPLE SIZE

The claim that the study had sample size of  $n = 10$  countries is incorrect.<sup>2</sup> Each of the 16 regression models represented in Figure 4 included, on average, 1362 data points (range 771-3493) on 52 subnational units (range 27-129). Each panel regression is, in effect, a ‘mini-meta-analysis’: the effect size is evaluated within each subnational unit, and the overall effect size is estimated from a pooling of these ‘within’ effects. So, while we aggregated the results to 10 countries, the sample size is not  $n = 10$ .

## 2 | COUNTRY SELECTION

In contrast to the suggestions in the letters, we left countries such as Denmark and New Zealand out not because they would demonstrate patterns that would support the role of restrictive measures, but exactly because the paltry spread of COVID-19 in these countries would prevent learning anything meaningful about the role of NPIs in these countries.<sup>2</sup> We anticipated separability challenges in identifying any nonpharmaceutical intervention's (NPI) effect when analysing changes in growth rates from a baseline of little (or, in many subnational units, no) growth. In our regression framework, no NPI would likely show a meaningful effect in Denmark or New Zealand where case growth rate was low, both before and after NPIs were implemented (hence no meaningful changes in case growth). We could have noted *ex ante* that we selected countries with more meaningful virus spread, though we do note that ‘Additional countries could provide more evidence, especially countries that had meaningful epidemic penetration’.

For countries like New Zealand with tepid epidemic growth, it is statistically and intuitively apparent that teasing apart which, among all the measures implemented, had worked is impossible. Viral entry and spread in New Zealand was limited relative to the United States, and more amenable to control. Less restrictive NPIs may well have been sufficient to maintain epidemic control. Some suggest that New Zealand's effective control can be ascribed to its highly restrictive lockdowns.<sup>5</sup> That opinion, unfortunately, has no evidence to support it beyond the anecdotal. As of March 2021, the highest death rates globally have occurred in countries that used prolonged and very restrictive measures, while the lowest death rates occurred in countries with more diverse responses. This is of course no proof of the futility of lockdowns, but it does call into question any claims of a much-worse counterfactual with less restrictive measures.

Experience from past pandemics has shown vast differences in disease spread across different locations, irrespective of measures taken, and we are seeing the same variability with COVID-19. Ignoring these plain-to-see epidemiologic patterns is a disservice to public health and society.

In terms of country selection, our analysis examines 4 countries more than the analysis after which it is fashioned (1.7 $\times$ ) and 2 countries less than a prominent, but problematic, modelling study of lockdown effects (0.8 $\times$ ).<sup>6-9</sup> We invite any researchers to add countries to our analysis: we made all the code available in large part for this exact reason.

## 3 | CONCEPTUAL SETUP

The comparisons we use are not ‘arbitrary’.<sup>2</sup> The distinction we made between more and less restrictive NPIs is a meaningful one, and it characterizes the study countries well. Idiosyncratic differences in implementation of various NPIs exist in all analyses evaluating the effectiveness of NPIs, including those that find favourable results for more restrictive measures.

Fuchs worries about omitting the period of declining daily case numbers, but this is a misunderstanding.<sup>4</sup> We measure growth of cumulative cases, which are monotonically increasing, and therefore never go below 0 (negative growth) in our study's figure 1 ("Growth rate in cases for study countries"). The data that we include cover the period up to the elimination of rapid growth in the first wave (Figure 1).

## 4 | STATISTICAL ISSUES

We would like to provide additional (hopefully clarifying) detail related to our statistical models. The binary  $Policy_{pcit}$  variables represent a set of indicators, each identifying an NPI (one for each NPI implemented in each subnational unit in each country), which changes value from 0 to 1 after the policy is implemented and for the policy's duration. The variable construction is equivalent to interacting the policy indicator with a 'post' time variable for each policy.<sup>3</sup> We implement panel regression models where the coefficients on the  $Policy_{pcit}$  variables identify 'breaks' in case growth patterns in each subnational unit following the implementation of each NPI identified by the specific  $Policy_{pcit}$  variables, rather than a difference-in-difference as suggested by Chini.<sup>3</sup> We analyse every pair of countries with more and less restrictive measures. For each regression, we include a full set of district-level binary dummy variables ( $\theta_{0,ct}$ ) for both countries in each model. These dummies remove the influence of time-invariant differences between districts and countries on the overall effect size. In addition, we also include day-of-week binary dummies  $\delta_{ct}$  that are unique to each country.

Rather than impose strong assumptions about the similarity between countries, our approach is to examine every pairwise comparison of countries that have adopted more or less restrictive measures. Thus, each regression model includes two countries—one with and one without more restrictive measures—each represented by the 'c' index that Chini found objectionable.<sup>3</sup> We do not pass a strong verdict on the role of parallel trends assumptions for causal identification here, but note that if it were indeed critical, that would invalidate most assessments of NPI effects that use similar econometric approaches, since the baseline trends are unique and highly nonlinear in each subnational unit.<sup>6</sup>

We note that, by mistake, we cumulated the case counts for the Netherlands twice. Correcting this, the trend for the Netherlands points more strongly to *enhanced* case spread with more restrictive measures (0.08 (0.00-0.17) vs Sweden and 0.13 (-0.11-0.37) vs South Korea).

We share concerns for confounding in analyses of observational data. However, possible confounding cuts in both directions. For example, in 9 out of 10 countries, we find a reduction in case spread following 'all NPIs', but even this beneficial effect could be due to confounding. Similarly, one may speculate

whether our finding of a negative effect size in Iran could be accounted for by inclusion of an 'authoritarian regime' variable.<sup>10</sup>

The issue of lags and timing seems to be a common concern, but makes no difference to the findings or interpretation. The timing of each NPI in each subnational unit of each country is explicitly modelled in the  $Policy_{pcit}$  variables. Introducing a lag does not alter any of the principal findings: those findings are driven by the similarity in the growth patterns of case counts between the compared countries, irrespective of any lags. We invite the letter-writers (and others) to introduce sensible lags into the statistical models and re-run the analyses (code and data are publicly available).

Fuchs' discussion of the effect size is incorrect. The lowest bound of the most favourable comparison for restrictive NPIs (Iran vs. South Korea) is  $-0.28$  (about 25% reduction over the entire period of the first wave when converted back from natural log scale; estimates from the study's figure 4). That is, the data suggest it is not very implausible that Iran's restrictive measures account for an additional 25% reduction in case growth over the entire period in comparison with South Korea. The data also suggest it is not very implausible that more restrictive measures account for upwards of 40% *increase* in case growth over the period in the United States, Spain and England. The negative extreme bounds should be emphasized no more than the positive extreme bounds.

## 5 | ONGOING CONCERNS WITH NPI EVALUATIONS

An underlying theme in the letters is that COVID-19's epidemic trajectories have been difficult to characterize, and have traced trajectories that often seem disconnected from the policies aimed at modifying these trajectories.<sup>11</sup> This introduces large uncertainties into any assessment of NPI effects. A recent modelling study illustrates the oscillatory epidemic dynamics under adaptive behaviour changes, underscoring the deeply endogenous relationship between behaviours and epidemic waves.<sup>12</sup> Policies, in that scheme, have a modulating impact of unknown extent on the realized epidemic patterns. The past year has revealed puzzling patterns of epidemic dynamics that have defied models that attribute much epidemic control to policies.<sup>13</sup> At the time of this writing, cases and deaths are declining across most US locations, despite models' predictions to the contrary.<sup>14</sup>

This points to a more generalized and pernicious challenge: how should NPI effects be studied? Simulation models are clearly problematic because their results are a direct function of input assumptions. Observational studies, especially using causal inference methods, have advantages. However, when the underlying dynamics are nonlinear and the policies are deeply endogenous, as in this case, attribution is precarious. This limitation is shared by all observational assessments

of NPI effects.<sup>6,15,16</sup> Randomization has been increasingly used for assessing the impact of real-world policies, and the value of knowing the benefits of NPIs, especially those with large health and welfare costs, would be enormous.<sup>17</sup> In all, we maintain that the science plausibly supports beneficial, null or harmful impacts on epidemic outcomes of highly restrictive measures, such as mandatory stay-at-home and business closures. Given their many uncontested harms to health and society, we believe that the extant literature does not provide strong support for their effectiveness at reducing case spread, and should be subjected to careful, critical and rigorous evaluation. If the benefits of such measures are negligible (or worse), their perpetuation may be, on balance, detrimental to the health of the public.

### CONFLICTS OF INTEREST

None.

Eran Bendavid<sup>1,2</sup> 

Christopher Oh<sup>1</sup>

Jay Bhattacharya<sup>2</sup>

John P.A. Ioannidis<sup>1,3,4,5,6</sup> 

<sup>1</sup>Department of Medicine, Stanford University  
Stanford, CA, USA

<sup>2</sup>Center for Health Policy and the Center for Primary  
Care and Outcomes Research, Stanford University,  
Stanford, CA, USA

<sup>3</sup>Department of Epidemiology and Population Health,  
Stanford University, Stanford, CA, USA

<sup>4</sup>Department of Biomedical Data Science, Stanford  
University, Stanford, CA, USA

<sup>5</sup>Department of Statistics, Stanford University,  
Stanford, CA, USA

<sup>6</sup>Meta-Research Innovation Center at Stanford  
(METRICS), Stanford University, Stanford, CA, USA

### Correspondence

Eran Bendavid, Department of Medicine, Stanford  
University, Stanford CA, USA.

Email: ebd@stanford.edu

### ORCID

Eran Bendavid  <https://orcid.org/0000-0002-8364-4711>

John P.A. Ioannidis  <https://orcid.org/0000-0003-3118-6859>

### REFERENCES

- Bendavid E, Oh C, Bhattacharya J, Ioannidis JPA. Assessing mandatory stay-at-home and business closure effects on the spread of COVID-19. *Eur J Clin Invest.* 2021;51:e13484. <https://doi.org/10.1111/eci.13484>
- Besançon L, Meyerowitz-Katz G, Flahault A. Sample size, timing, and other confounding factors: toward a fair assessment of stay-at home orders. *Eur J Clin Invest.* 2021;12:e13518. <https://doi.org/10.1111/eci.13518>
- Chini EZ. Letter to Editor. *Eur J Clin Invest.* 2021;e13556. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/eci.13556>
- Fuchs H. Comment on Bendavid E, Oh Ch, Bhattacharya J, Ioannidis JPA Assessing mandatory stay-at-home and business closure effects on the spread of covid-19. *Eur J Clin Invest.* 2021;4:e13529. <https://doi.org/10.1111/ECI.13484>
- Cousins S. New Zealand eliminates COVID-19. *Lancet.* 2020;395(10235):1474. [https://doi.org/10.1016/S0140-6736\(20\)31097-7](https://doi.org/10.1016/S0140-6736(20)31097-7)
- Hsiang S, Allen D, Annan-Phan S, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature.* 2020;584(7820):262-267. <https://doi.org/10.1038/s41586-020-2404-8>
- 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>
- Soltész K, Gustafsson F, Timpka T, et al. The effect of interventions on COVID-19. *Nature.* 2020;588(7839):E26-E28. <https://doi.org/10.1038/s41586-020-3025-y>
- Bryant P, Elofsson A. The limits of estimating COVID-19 intervention effects using Bayesian models. *medRxiv.* 2020;2020:20175240. <https://doi.org/10.1101/2020.08.14.20175240>
- Democracy Index 2020. Economist Intelligence Unit. Accessed March 1, 2021. <https://www.eiu.com/n/campaigns/democracy-index-2020/>
- America's two largest states are fighting covid-19 differently. The Economist. Published online February 6, 2021. Accessed March 1, 2021. <https://www.economist.com/united-states/2021/02/06/americas-two-largest-states-are-fighting-covid-19-differently>
- Arthur RF, Jones JH, Bonds MH, Ram Y, Feldman MW. Adaptive social contact rates induce complex dynamics during epidemics. *PLoS Comput Biol.* 2021;17(2):e1008639. <https://doi.org/10.1371/journal.pcbi.1008639>
- Marchant R, Samia NI, Rosen O, Tanner MA, Cripps S. Learning as We Go: An Examination of the Statistical Accuracy of COVID19 Daily Death Count Predictions. arXiv:200404734 [q-bio, stat]. Published online May 24, 2020. Accessed February 21, 2021. <http://arxiv.org/abs/2004.04734>
- Conlen M, Lu D, Glanz J. Why Vaccines Alone Will Not End the Pandemic. The New York Times. <https://www.nytimes.com/interactive/2021/01/24/us/covid-vaccine-rollout.html>. Published January 24, 2021. Accessed February 21, 2021.
- Chernozhukov V, Kasahara H, Schrimpf P. Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. *J Econom.* 2021;220(1):23-62. <https://doi.org/10.1016/j.jeconom.2020.09.003>
- Brauner JM, Mindermann S, Sharma M, et al. Inferring the effectiveness of government interventions against COVID-19. *Science.* 2021;371(6531): <https://doi.org/10.1126/science.abd9338>
- Cristea IA, Naudet F, Ioannidis JPA. Preserving equipoise and performing randomised trials for COVID-19 social distancing interventions. *Epidemiol Psychiatr Sci.* 2020;29:e184. <https://doi.org/10.1017/S2045796020000992>

**How to cite this article:** Bendavid E, Oh C, Bhattacharya J, Ioannidis JPA. Authors Response to Letters to the editor regarding: 'Assessing mandatory stay- At- Home and business closure effects on the spread of COVID- 19'. *Eur J Clin Invest.* 2021;51:e13553. <https://doi.org/10.1111/eci.13553>