

July 06, 2020

Dr. Elizabeth Loder

The BMJ

RE: Effectiveness of physical distancing interventions on COVID-19 incidence: natural experiment in 149 countries

(Manuscript Ref. BMJ-2020-059328)

Dear Dr. Loder,

Thank you very much for the opportunity to review our paper “**Effectiveness of physical distancing interventions on COVID-19 incidence: natural experiment**” based on the comments and feedback from the Editorial committee and the external reviewers. We believe the revised version has improved substantially. We sincerely thank the Editorial team and the reviewers for such a comprehensive critical appraisal of our paper.

Please find below our point-by-point responses to the comments from the Editorial committee and the reviewers, along with the corresponding changes in the manuscript, as applicable. We made every effort to adhere to the BMJ guidelines in revising the manuscript, but we will be pleased to provide any additional information, and/or make any additional changes, as needed.

Thank you very much once again for considering our manuscript for publication in The BMJ.

Sincere regards,



Nazrul Islam, MBBS, MSc, MPH, PhD

Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU)

Nuffield Department of Population Health, Big Data Institute

University of Oxford, Oxford, UK

Email: nazrul.islam@ndph.ox.ac.uk

Co-directors: Sir Rory Collins FRS FMed Sci BHF Professor of Medicine & Epidemiology

Sir Richard Peto FRS FMed Sci Professor of Medical Statistics & Epidemiology

Deputy Director: Colin Baigent FFPH FRCP Professor of Epidemiology

Jane Armitage FFPH FRCP
Professor of Clinical Trials & Epidemiology

Sarah C Darby PhD
Professor of Medical Statistics

Michael Lay DPhil
Head of Project Information
Science

Louise Bowman MRCP MD
Associate Professor of Clinical Trials
& Epidemiology

Christina Davies BMBCh MSc
Senior Research Fellow

Christine Marsden PhD
Director of Finance & Administration

Zhengming Chen MBBS DPhil
Professor of Epidemiology

Richard Gray MA MSc
Professor of Medical Statistics

Sarah Parish DPhil
Professor of Medical Statistics
& Epidemiology

Michael Clarke DPhil
Professor of Clinical Epidemiology &
Public Health Medicine

Michael Hill DPhil
Laboratory Scientific Director

Tim Sprosen DPhil
Head of Regulation &
Knowledge Transfer

Robert Clarke FFPH FRCP
Professor of Epidemiology & Public
Health Medicine

Martin Landray PhD FRCP
Professor of Medicine & Epidemiology

Alan Young DPhil
Director of Information Science

Comments from the The BMJ's manuscript committee meeting

* We agree with reviewers that this is an important look at the impact of policies that have had a major effect on people's lives and livelihood. We agree, too, that it is difficult to study these interventions but feel you have done the best you can with the imperfect data available. We are pleased that you prominently acknowledge the limitations of the data and methods. Examining the associations between these 5 physical distancing measures and actual COVID case rate is useful, since many of our public health interventions so far have been guided by modeling studies.

Authors' response: We thank you very much for your kind appreciation of our work.

* We think the biggest limitation of the data is that it does not take into account the heterogeneity of the response within a country, i.e, all states in the U.S. are treated the same.

Authors' response: **DONE.** We agree that this is one of the limitations of our analysis. Unfortunately, more granular data at state/province (or smaller) levels are not yet available on physical distancing policies. As you suggested below, we have **conducted a sensitivity analysis excluding seven larger countries** (Brazil, Canada, China, India, Russia, United Kingdom, and United States) and found that the primary finding is robust to this sensitivity analysis. We added the results from this sensitivity analysis in the Results section, and added the details in the Supplementary Appendix. Further, we added this as one of the limitations in the Discussion section.

Nor do you seem to take into consideration the "lifting" of measures, which is now taking place.

Authors' response: **DONE.** We agree that some countries have recently started to "lifting" the physical distancing measures. Data on this is not yet available systematically across the countries. However, to address this issue, **we restricted our revised analysis up to May 30, 2020, or 30-days since the implementation of first intervention, whichever came first.** Hope, this addressed this issue.

Are you willing to update your analyses at least through May 30?

That will provide another 6 weeks of data on top of the 10 you are using, since there were not many infections during the first 2 weeks of January.

Authors' response: **DONE.** Thank you for this suggestion. As suggested, we **updated the analysis with the data up to May 30, 2020**, which added additional 30 countries to a total of 149 countries. However, in the regression model, the post-intervention follow-up was restricted to May 30, 2020, or 30-days post-intervention.

The reason for restricting our revised analysis to 30-days post-intervention is to avoid including longer post-intervention time which may 'unequivocally prove the efficacy', as also mentioned in the Editorial comment below. This approach also minimises dramatic changes in other covariates that may compromise the assumption of ITS analysis, as also suggested below in one of the comments from the Editorial team.

* Might you also share your data for others to examine?

Authors' response: **DONE.** Yes, we will make all the data used in our analysis available in a dedicated GitHub page. The GitHub link has been provided in the revised manuscript; the data will be made available upon publication of the manuscript.

* One editor commented that perhaps this is a less pressing question now that many countries are reopening. He mentioned that "The difficult part is how to reopen without getting things out of control again. We are now seeing resurgence in Europe and the US." Do you have any ideas about which of these measures might be important if partial lockdowns are needed? You might briefly discuss this.

Authors' response: **DONE.** We agree that a pressing question is which interventions to keep as some countries started to ease the lockdown. As suggested, we **discussed this in the discussion section**. Our results suggest a combination of at least three interventions (school closure, workplace closure, and restriction on public gatherings) so that the public transport may be kept open for people working in the key service sectors. Any intervention combination without restriction on public gathering may not be helpful in reducing the COVID-19 burden.

* However, a US editor thought this remains an important question. "Many in the US (perhaps associated with political affiliation) have questioned the effectiveness of protective measures, or that the costs outweigh the risks. Seems I can't open Twitter without some such post. Thus, I do think the RQ remains timely."

Authors' response: We agree with the Editor that we need robust evidence on the effectiveness of physical distancing as it is downplayed by many political leaders across the world.

* Is there a reason why you did not consider use of face masks? One editor remarked that "it would be nice to have something on face masks. It's one of the most actionable items and is a political fault line. We are seeing places with higher than ever rates and people debating whether the city should be allowed to have regulations requiring face masks."

Authors' response: **GLOBAL DATA UNAVAILABLE.** We agree that use of face coverings is likely one of the potentially effective, and most actionable, public health interventions in reducing the viral transmission. However, robust data on the use of face coverings across the globe is not yet available. We discuss this in our Discussion section.

* The methods are somewhat difficult to follow. You calculate the slope of the infection rate before and after the introduction of restrictive measures per country; this assumes that rate of testing stayed the same during that period. We think that is a questionable assumption. Might you comment?

Authors' response: **DONE.** We agree, and therefore, **restricted our analysis up to 30-days post-intervention to minimise** any substantial change in the testing rate. We further discuss the paucity of valid longitudinal data on COVID-19 testing in our discussion.

Also, even without any measures, one would expect the rate of new infections to drop off at a certain point, so it then depends on where you take the trajectories in the curve from which you calculate the IRRs. Thus we wondered whether this unequivocally proves the efficacy of the measures.

Authors' response: **DONE.** We agree that inclusion of data further along the epidemic curve for each country could potentially more unequivocally prove the efficacy of the measures. However, as

indicated by the editors, past the peak of the epidemic curve incidence is declining, sometimes rapidly and substantially. We, therefore, **restricted our analysis up to 30th May 2020 or 30-days post-intervention, whichever came first**. In practice, this ensured that assessment of effectiveness largely covered the upswing on the epidemic curve in each country.

* Related to this, we wondered about the data from some countries in particular. Most countries show a lot of variability with the dots all over the place. A couple of countries have data that fits almost too well on curves. Is this a signal about the validity of the data? For example, China has a well behaved epidemic curve going smoothly up and almost symmetrically and smoothly down...

Authors' response: Scatter of data is, to an extent, the product of the size of the data set (in general related to the size of the country's population). However, it may not be surprising to find that some countries manipulated their data for political reasons, but it is difficult to convincingly prove this without empirical evidence. Since this phenomenon was not observed in most of the countries analysed in our study, this may not have a substantial effect on the overall findings of this study. Also, a sensitivity analysis that excluded China along with other larger countries did not show any substantially different result from the primary analysis.

* We think an exponent is missing in your equation.

Authors' response: **DONE**. Thank you, and apologies; it has been corrected in the revised version.

* The approach you have taken to individual country policies is to broadly categorise and assume that with enough data points it will all average out. What then does the average mean for a specific policy recommendation since every one of these can vary widely in how they are implemented? (1 meter or 2, indoors or out, families included or not etc). Please discuss the heterogeneity of the effect within country.

All the overall US effects were negative, and we think that's because of the difference across states and how guidance was implemented. Can you address this in some way? While you may not be able to do it at the state level in the US, maybe do an analysis without the US in it?

Authors' response: **DONE**. Thank you for the suggestion. We have **discussed the issue of within-country heterogeneity**, compliance, and adherence to proposed policy etc., and acknowledged that future studies with more granular data may provide further insights on this. We also conducted a **sensitivity analysis excluding the larger countries** (including the US), as suggested.

* Our statistician commented that the interrupted time series does allow for a 'natural experiment', but she felt this may not warrant the level of causality inferred in the text. She noted that reviewer May review gives some good pointers for toning down the conclusions, which we would like you to do.

Authors' response: **DONE**. We thank the statistician for this helpful comment. As you will appreciate, we have made every effort to highlight the limitations of the study, and not to overstate our results. **We revised the manuscript to further tone down any implied note on causality.**

* Our statistician agreed with the reviewers who question the 7-day time lag and she also wonders whether other lags should have been investigated. Please consider this.

Authors' response: **DONE.** We thank the statistician and the reviewer for this helpful feedback. We used a 7-day time lag as this was used in a previous study, but examined the robustness by using **two additional analyses with a 5-day and 10-day lagged time-frame.** The findings from these two analyses were similar to those from our primary analysis. We have reported these in the revised version as a sensitivity analysis.

* There are some (relatively small) discrepancies between the values given in the abstract, the text and the supporting appendices (eg. In relation to the order of interventions).

Authors' response: **DONE.** Apologies for this; we checked the numbers in the revised version carefully.

* The outcome variable for figure S7 needs labelling/explanation.

Authors' response: **DONE.** Apologies again. The PDF conversion had some issues in the earlier version. The revised manuscript has the Y-axis clearly labelled.

* Our statistician commented that "The data is well presented as individual country plots of incidence, policy implementation timeline, and then fitted models with raw data."

Authors' response: **DONE.** Thank you for the suggestion. We reorganised the figures as suggested.

In your response please provide, point by point, your replies to the comments made by the reviewers and the editors, explaining how and where you have dealt with them in the paper.

Comments from Reviewers

Reviewer: 1

Comments:

- Are the questions the paper addresses relevant and important to patients and/or carers? Yes - this is highly relevant to patients

Authors' response: Thank you.

- Are there topics or issues that are missing, or need to be highlighted more?

I'm not sure how many 'lay' people will be reading this article, but I wonder if a sentence in the abstract needs to be re-worded to be clearer - "the reduction in COVID-19 incidence was comparable with and without public transport closure when the other four interventions were implemented" - I'm not sure whether the incidence being comparable, means that the incidence was the same whether public transport closures happened or not - so public transport closures made no difference to the incidence? I really understood what you meant in the conclusion of the abstract - it is clearly explained here. Also, another sentence in the abstract - "compared with delayed implementation

after school and workplace closure" - do you mean compared with delayed implementation OF school and workplace closure? I think it is very important that the abstract can be widely understood because if lay people do pick up this article - they will most likely read the abstract for an easy to digest summary.

[Authors' response: DONE.](#) We thank you for this helpful comment. We revised the abstract accordingly. We hope this version reads better.

- Is the treatment or intervention suggested or guidance given something which patients/carers can readily take up? or does it present challenges?

Yes - this provides clear data on the degree of success of social distancing measures and the importance of adhering to them to 'flatten the curve'.

However, what is not clear to the public is how comparable this data is on social distancing to other measures - such as face masks. If social distancing has reduced the incidence by 13% how does this compare to other interventions (like face masks)? The interventions are listed, but comparisons are not provided. It may not be possible to state this yet if figures are not available?

[Authors' response: GLOBAL DATA UNAVAILABLE.](#) We agree with the reviewer that additional data on face coverings would be helpful. However, as the reviewer noted, robust data on the use of face coverings is yet to be available. Therefore, we could not examine the effect of face coverings in our paper.

- Are the outcomes described/measured in the study important to patients/carers? Are there others that should have been considered?

Yes - the outcomes are highly relevant

[Authors' response: Thank you.](#)

- Do you have any suggestions that might help the author(s) strengthen their paper and make it more useful for doctors to share and discuss with patients/ carers?

1. the changes to the abstract that I have suggested

[Authors' response: DONE.](#) Thank you. We revised the abstract for further clarity.

2. perhaps a little more reference to psychological literature - early implementation of preventing public gatherings and 'lockdown' are the most effective, but for healthcare staff to encourage their patients to adhere/comply to this guidance perhaps it would help for them to be able to offer evidence based strategies to 'cope' with this. Perhaps there will be 'lock-down fatigue' at a future point in time. If doctors could say - 'this is the evidence that lockdown is effective - and here are evidence based suggestions on how we can cope with this together' - perhaps the information might be more 'well - received'? The article probably only needs a few references in the discussion to show that this has been thought about?

[Authors' response: DONE.](#) Thank you for this helpful suggestion. We added a separate paragraph in our discussion section in light of recent publications.

- Do you think the level of patient/carer involvement in the study could have been improved? If there was none do you have ideas on how they might have done so?

I don't think patient involvement is needed to produce a paper of this kind. However, there will be merit in un-picking/un-ravelling patients response to this document. We are talking about behaviour change and restrictions on how people live their lives, so providing an obvious channel for the public to respond to this document could be helpful to explain how academics/scientists/policy makers can foster better relationships with patients and the public to carry forward recommendations into practice.

[Authors' response:](#) We completely agree with this. We believe BMJ will serve as a great platform for public discussion on this. We will do our best to communicate these findings to as many stakeholders.

Reviewer: 2

Comments:

This study is a natural experiment and meta-analysis with a robust design to evaluate the impact of the implementation of physical distancing measures (school closures, workplace closures, public event cancellation, public transport closure and 'lockdown' measures) on the trajectory of the global pandemic of COVID-19. One hundred nineteen countries/regions are included between January 1 and April 16, 2020. The following are my feedback on this paper:

- Originality:

v This study attempts to fill up the knowledge gap on the impact of physical distancing policies using a cross-country, comparative approach.

- Importance of work to general readers:

v Some of the findings may support policy decisions, taking into considerations the limitations of the findings. The findings can only be taken as reference and should be used with caution. Such findings may not be applicable to every country because of diverse characteristics and pandemic situations.

[Authors' response:](#) Thank you very much for your kind appreciation of our work. We agree with the cautionary notes, as we also highlighted many of the potential limitations in our discussion.

- Scientific reliability:

v The background of the research team is strong, consisting of experts in epidemiology, public health and statistics from distinguished tertiary institutions.

v Data on the policy interventions was derived from various reliable sources.

- Research question:

v The general aim of this study instead of research questions is stated (p.6, lines 27–28).

[Authors' response:](#) **DONE.** Thank you for your comments. We have added the research question in the revised version.

v The findings of the study have generally achieved the aim of the study.

- Overall design of the study:

v This study used natural experiment and meta-analysis, adopting an interrupted time-series design.

- Study participants:

v The data of participants were anonymous and aggregated without any personal information.

- Methods:

v The criteria to be included in the analysis were reasonable.

v Although the authors hypothesised a '7 day lag time' (p.8, line 53) to justify 'the use of the first 7 days since the implementation of the intervention as the pre-intervention period', why they did not use 7 days 'before' the implementation of the measure(s) as the baseline?

Authors' response: **DONE.** Apologies for the confusion in the wording. We have considered any time before the intervention *plus* the first seven days of intervention as pre-intervention time. This has been further clarified in the Methods section.

On this note, we have also conducted two sensitivity analyses using a 5-day and 10-day lagged time-frame to examine the robustness of our primary result. The results were similar to those from the primary analysis. We have added these results in the manuscript (and details in the Supplementary Appendix).

v Ethics approval was waived because all the data were anonymous and publicly available.

v A number of country-level characteristics, which may affect the policy intervention and the incidence of COVID-19, were assessed in meta-regression by the authors (p.9, lines 42–60). However, a recent systematic review indicated that the use of face mask could result in a large reduction in the risk of infection (Chu et al. 2020). Therefore, whether or not countries implement the mandatory mask policy in public areas may have an impact and may be considered (see reference below).

Authors' response: **GLOBAL DATA UNAVAILABLE.** Thank you for this suggestion. We indeed wanted to examine the effect of use of face coverings in the meta-regression. However, global data on face covering use is not yet available. The Lancet paper included only a few studies on face coverings from select countries (with an overall 'low' certainty). The Aljazeera news has some data on a few countries, but we hope you will agree that this list is not comprehensive (i.e., not systematically collected to be used for research purposes). The list does not have data on the extent of use by countries to be used in the meta-regression. We have highlighted this in our discussion as one of the limitations of the data, and called for more robust data for future studies. **We have cited the Lancet paper on face coverings as suggested.**

- Results:

v Appropriate data analyses were adopted, including but not limited to the use of interrupted time series analysis of each country's data to model the population incidence of COVID-19 over time, random-effects meta-analysis and sensitivity analyses.

v The statement 'regions that took earlier and aggressive physical distancing measures grew faster (in terms of what?) in the post-pandemic period' needs to be elaborated (p.16, line 6).

Authors' response: **DONE.** Thank you for pointing this out. The word 'economically' was implied, but we have clarified it in the revised version.

v In general, presentable study findings with relevant tables/figures were provided.

- Interpretation and conclusions:

v More discussions could be made to some interesting findings arising from this study, for example:

- Why public transport closure measure had little effect when the other four physical distancing measures were in place? Would it be due to that the other measures such as school closures, workplace closures and public event cancellation already decrease the number of people using public transport and resulted in no additional benefit of this measure on the IRR incidence?

Authors' response: **DONE**. We agree, and elaborated further on this in the revised version. Thank you.

- Why higher percentage of population $\geq 65\%$ was associated with greater reduction in the pooled IRR (p.12, line 17, p.15, line 2) and not the other way round?

Authors' response: In line with the ecological nature of the analysis, countries with a higher proportion of elderly population are more likely to be more wealthy with many additional social capital, such as higher GDP (which also had similar effect), better healthcare, income insurance etc. Elderly populations are also likely to adhere to physical distancing and less likely to be involved with jobs/activities with high exposure probability compared to their younger counterparts.

- Some discussions on the interventions implemented within a 7 day span could be made, as a number of countries 'favour status quo' instead of 'physical distancing measures'. Were there any possible reasons? (pp. 50–55).

Authors' response: **NOT APPLICABLE** in the revised version. While there could be some heterogeneity in the effectiveness across the countries, in the revised analysis combinations and sequence analysis (within a 7-day time frame) shows that the 95% CI of all the countries with the point estimate favouring "status quo" include the null value.

v Conclusion could generally summarise the focus of the study.

Authors' response: **DONE**. Thank you. We further revised the Discussion section, and added the Summary Box to highlight the major points.

- References:

v Recent and relevant references were included to support arguments.

v The following references can be considered:

⌘ Chu DK et al. (2020). Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet. Date of Electronic Publication: 2020 June 01.

Authors' response: **DONE**. Thank you for the suggestion. We have added this reference in the discussion section.

⌘ Which countries have made wearing face masks compulsory?

<https://www.aljazeera.com/news/2020/04/countries-wearing-face-masks-compulsory-200423094510867.html>

- Abstract/summary/key messages/what this paper adds:

v The summary was succinct and generally covered the main points of the study.

Authors' response: Thank you.

Reviewer: 3

Comments:

Thank you for inviting me to review this paper on the effects of five key distancing interventions on the COVID-19 pandemic.

It presents actual data on the numbers of COVID-19 cases and reported deaths from 119 countries after the implementation of five key distancing interventions – school and workplace closures, event cancellation, “lockdown measures” and closure of public transport. All five were introduced in 77 countries. The team have looked at what measures were adopted, and how soon after the first reported cases, and how the numbers of cases and deaths rose before and after the measures were implemented.

Its strength lies in the presentation of actual data rather than from mathematical modelling or extrapolation from outbreaks caused by MERS or SARS, both of which are probably rather less transmissible.

The study design is appropriate. It is well written and presented, with a clear message. There will be minimal work for a copy editor to do.

It certainly belongs in a general journal such as the BMJ, and I imagine that will be much cited by other authors, and will quite probably also attract international media attention. The conclusions will be helpful to policymakers, not least when some UK politicians are openly questioning the value of lockdown measures.

This can't have been easy to do, and inevitably the data won't have been uniformly robust, however they discuss the limitations appropriately. They were not able to drill down to more specific measures such as the widespread wearing of masks, the use of 'phone apps to trace contacts, track and trace systems, differences in availability and use of PPE, the numbers of available beds and ventilators, all of which might have had an effect on both transmission and clinical outcome.

Albeit on a global scale, this demonstrates just how difficult it is to undertake studies on the efficacy of any infection control measure, as the transmission of any outbreak is always multi-factorial and will also be affected by things that you can't possibly control for, including human behaviour and compliance, or perhaps how rigorously individual state authorities were prepared to implement their various control measures.

The authors have done a good job in distilling such a large amount of complex data into a manageable document with a couple of short key messages, namely:

1. If the four measures (school and workplace closures, public event cancellation and “lockdown” measures) were already in place, then shutting down public transport closure didn't confer any extra benefit.
2. The measures worked best when public events cancellation and lockdown were implemented first.

The first conclusion may not be so very surprising –if only key workers can go to work; you can't go to school, or a public event and are not supposed to leave your house without very good reason, apart from a visit to the supermarket perhaps, where is there left to travel to? Trains, trams, tubes and

buses should have been much emptier anyway. Nonetheless, that's an important consideration for policymakers.

At 300 pages, it is the longest submitted manuscript that I have ever seen either as a journal EIC or reviewer. The online supplementary material is interesting, but there is an awful lot of it, although nothing that I would excise.

That said, I wonder whether the individual figures in S5 and S6 might be better presented together for each country rather than each being listed sequentially? It would be interesting to be able to view the dates the measures were introduced, cases and deaths (s5) side by side with the modelling data (s6).

Authors' response: DONE. Thank you very much indeed for your appreciative words. As you highlighted, some of the other potential variables could not be addressed due to unavailability of these data. We aimed to highlight these limitations to the best of our ability. As suggested, we have re-organised the figures sequentially by country in the revised version.

Reviewer: 4

Comments:

I have struggled mightily with my recommendation for this paper, as I simultaneously believe it provides extremely important preliminary evidence for the effectiveness of social distancing — support which is desperately needed as social distancing measures are challenged by those who wish to dismiss the dangers of COVID-19 as “fake news” — but also believe it relies on data that (through no fault of the authors) is fundamentally flawed.

The authors correctly point out that most data concerning the effectiveness of social distancing from covid-19 comes from modeling. The greatest strength of this study is its reliance not upon modeling, which is subject to potentially biasing input and algorithms, but on actual data.

While the use of some modeling techniques (e.g. interrupted time series analysis) remained necessary to, for example, establish “controls” specific for each country, the primary data input reflected actual testing results as collected in each country. Unfortunately, this is also the study's greatest weakness.

Due most directly to the failure to implement a coordinated, consistent testing strategy both globally and, in most cases, regionally or nationally, rates of positive testing results might well reflect changing testing practices rather than actual effects on incidence. This is true not only for “total diagnosed cases”, but also for positive result incidence rates. For example, in the U.S. early shortages of testing kits led covid-19 testing to be restricted in many areas to only those showing obvious overt symptoms, or known to have been exposed to others who tested positive. Once testing expanded beyond these individuals that we had strong independent reason to believe would test positive, we should naturally expect the incidence ratio of positive test results to lower. Because very little coordination — let alone supervision and consistency in application — of testing strategies occurred in the U.S., it is nearly impossible to know how we might account for variable testing practices in any analysis of positive result incidence. As illustrated by the early testing restrictions just described, there is little consistency in testing practices even within local testing sites, let alone between such sites. To this day, testing data in the U.S. remains a mess, subject to political manipulation resulting in the resignation of state health officials who refuse to “cook the results”, and with newspapers reporting that even CDC data is failing to properly sort antibody testing from active case testing, completely

corrupting the utility of data for scientific purposes. The WHO has faced similar accusations of subverting epidemiology to political pressures internationally, further calling into question the accuracy and utility of data collected globally.

Perhaps the strongest scientific conclusion to be drawn from this work (in terms of definitive knowledge), then, is the illustration of what it might have been possible to know had better coordination of testing taken place. While data supporting the effectiveness of social distancing is extremely important, maintaining trust between the public and scientific advisors is even more important. We must be careful, then, to not mislead or overplay convenient findings, but instead acknowledge the limitations of what conclusions we can draw. Only by acknowledging our failures in systematic testing and data collection can we learn from our mistakes and avoid repeating these. Glossing over the flawed nature of data in order to support a desired conclusion risks violating the trust of the public. I want to be clear that these criticisms are not directed toward the study authors, who have done admirable work, but toward the shameful politicization of the global public health infrastructure, which has resulted in corrupted data.

Nonetheless, the study provides strong -if not definitive for the reasons described above—support for social distancing. The fact that effectiveness is maintained over so many different data collection mechanisms and locations (individual countries) is, to me, strongly suggestive of social distancing effectiveness. In addition, the examination of specific country-by-country social distancing strategies is —by itself— a valuable resource in pandemic response planning. In essence, I view this study's results as I would preliminary data for a grant: what data we have is strongly suggestive of a conclusion, but it lacks the quality, rigor and consistency needed to definitively rely on its conclusions (thus the need to do the study one is seeking a grant for). I suspect the study's conclusions are correct...but we cannot know this from data collected (by countries, not the authors!) so un-systematically and without care to detail.

In short, this data is very helpful and suggestive, but would be much more helpful were there better implementation of testing (specifically, consistent and standardized) such that more reliable conclusions might be drawn.

[Authors' response: DONE.](#) We thank the reviewer for such a comprehensive critical appraisal. We completely agree with these points. We highlighted many of these issues in our discussion section with the emphasis on need for more robust data, as the reviewer highlighted. We revised the manuscript thoroughly further to tone down to avoid pointing to definitive causality given the limitations of the study design and existing data.

Because I believe the study is of such great importance, because I think the authors did such a good job with the flawed data they have to work with, and because I believe the results are highly suggestive and conclusions probably correct (though not definitively proven due to unavoidable use of flawed data), I strongly recommend publication of this important work, but would also strongly suggest it be accompanied by an independent commentary that highlights the limitations of conclusions that are based on data that is flawed (through no fault of the authors: they have done the best that can be done with the flawed data available) — and how this emphasizes the need for greater coordination, and consistent, systematic implementation of both interventions and, perhaps most importantly, testing strategies so that legitimate conclusions may inform future response as well as refine current efforts. This limitation is of sufficient importance to merit its being highlighted in an accompanying commentary, rather than lost in the minutiae of other study limitations.

[Authors' response:](#) We agree with these invaluable comments. As much as we attempted to highlight the limitations of the data in our discussion, and made every effort to not overstate our findings, further commentary and discussions will be much appreciated. However, we leave this to the Editorial team for their consideration.

Reviewer: 5

Comments:

Thank you for kindly inviting me to peer review this paper by Dr Islam and colleagues, entitled "Effectiveness of physical distancing interventions on COVID-19 incidence". This study evaluated the impact of implementing physical distancing interventions on the incidence of COVID-19 utilising data from 119 countries.

There is certainly a lack of data that is currently available on the efficacy of physical distancing measures on COVID-19 infection rates and hence, this piece of work addresses an important issue that would be very useful for public healthcare systems across the globe.

In general, the article is well written and presented – the authors have clearly described the methods and analyses they have undertaken. The major strength of this study is that the authors have been able to rapidly combine and analyse data from 119 different countries.

Major comments

- Given the rapid nature of data collection and the complexities of collecting and collating data from 119 different countries, there are a number of intrinsic limitations to the study design. However, these are well described and acknowledged by the authors in the limitations section of the discussion.
- The abstract should provide more details on what major data sources the authors utilised to perform this study in the "design" subheading (i.e. mention the Oxford COVID-19 Government Response Tracker – which was the key resource that enabled the authors to perform this analysis).

[Authors' response:](#) **DONE.** Thank you for your suggestion. We have added this in the abstract in the revised version.

- It would be useful to provide more background details regarding the Oxford COVID-19 Government Response Tracker resource in the methods section i.e. what data is collected, how the data is collected, how the data is used etc.

[Authors' response:](#) **DONE.** Thank you for this suggestion. More details have been added to the revised version.

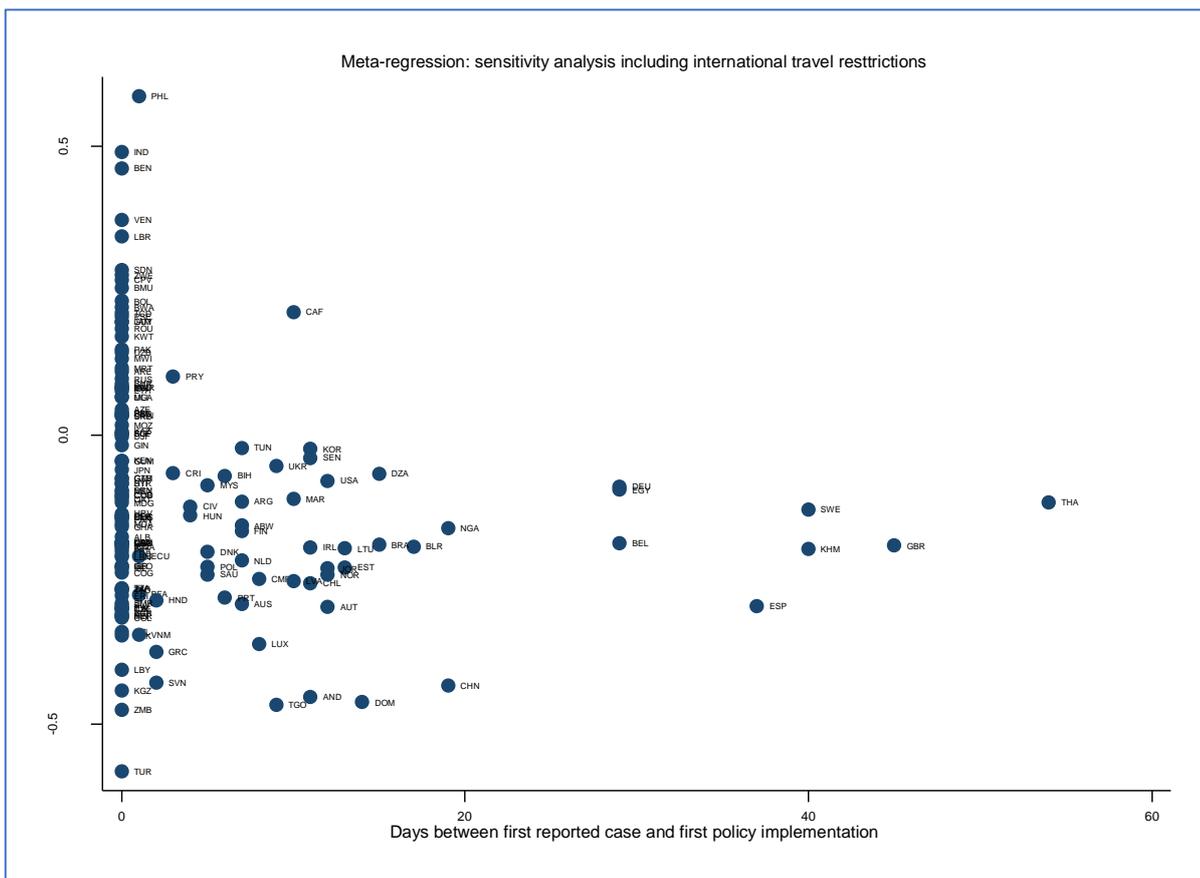
- The Oxford COVID-19 Government Response Tracker resource provides data on 8 containment measures – however only 5 were used for this study. Measures not utilised in this study but available in Oxford COVID-19 Government Response Tracker resource include international travel restrictions, restrictions on gatherings and stay at home restrictions. Why did the authors choose not to include these important physical distancing policies for their analysis? This is an important limitation of this

piece of work – if the authors are able to include these other measures in their analysis, then it would be pertinent to re-run the analysis with these measures.

Authors’ response: DONE. Our previous analysis included data up to April 16, 2020. The new variables (stay at home, and restriction on gathering) have been added on the 28th of April (Ref.: <https://www.bsg.ox.ac.uk/sites/default/files/OxCGRT.%20What%27s%20changed%2024%20April%202020.pdf>), and so we could not add those additional variables in our previous analysis.

However, as you suggested, we have included these two additional variables. The only variable that we did not include was restriction on international travel because, while this belongs in a viral containment strategy, it is not strictly a physical distancing measure. Moreover, international travel restrictions of one country often affect other countries regardless of whether those affected countries have implemented the same restrictions; this may violate the assumption of independence across the countries in the meta-analysis.

However, to respect your suggestion, we have checked this in a separate analysis including international travel. The results were similar, i.e., the effect was still not significant. Please see the figure that includes international travel below. The reason for that may be because many countries implemented international travel restriction at the very outset, but many of them did not do well in restricting the epidemic. Previous research showed that international travel restriction is ineffective in viral containment (Wells et al. PNAS, 2020; Yu and Keralis. HHR. 2020). However, we did tease out the effect of delay in lockdown (as opposed to delay in any physical distancing measure), and found that a delay in lockdown was associated with a slower reduction of COVID-19 incidence.



• On Page 12, Lines 10-14, the authors state “Meta-regression did not identify any effects on the IRR of days since the first reported case until the first policy implementation....”. This is quite a surprising

finding, and “contrary to anecdotal data from some countries that implemented these policies earlier” as the authors themselves state in the discussion. Could this result have been confounded by any other factor e.g. differences in testing rate during the period between first reported case and first policy implementation vs the period after first policy implementation? Alternatively, this finding may also be due to some countries implementing one of the 3 physical distancing measures which the authors did not utilise in the analysis, as their first implemented policy (for instance many countries implemented international travel restrictions well before other physical distancing measures: <https://www.brookings.edu/2020/04/02/the-early-days-of-a-global-pandemic-a-timeline-of-covid-19-spread-and-government-interventions/>)

Authors’ response: **DONE.** Thank you for your comments. As stated above, we have now included additional variables recently added to the database. We also checked the effects by including international travel, which was similar to our primary result.

To minimise the impact of substantial change over time in terms of COVID-19 testing, we have restricted our time-series model up to 30-days post-intervention. However, in our revised analysis, we teased out the effects of multiple treatments, and found that a delayed implementation of lockdown was indeed associated with a lower effect in reducing the COVID-19 incidence.

- On Page 13, lines 42-56, the authors state: “Hence, there was evidence of greater effectiveness when public event cancellation and population movement restrictions (lockdown) were implemented earlier, before school and workplace closures (pooled IRR: 0.80, 95% CI: 0.65-0.98, N=3) as opposed to when these were implemented later (i.e. after school and workplace closures) (pooled IRR: 0.89, 95% CI: 0.68-1.16, N=3).” The authors go onto state this as a main finding in the discussion on Page 14 (“A greater reduction in incidence was observed when public event cancellation and lockdown were implemented earlier together with school closure and workplace closure”) and also in the abstract. I have two concerns with this finding which I would like the authors to kindly address. Firstly, although the IRRs were mildly different (0.80 vs 0.89), the confidence intervals overlap with each other – therefore there may be no statistically significant difference between these two sequences of interventions. Secondly, this finding is based on data from just 3 countries for each sequence of interventions, and hence is greatly underpowered to be able to draw generalisable conclusions applicable for all countries.

Authors’ response: **DONE.** Thank you for your comments, and we agree that findings from fewer countries should be cautiously interpreted. In our revised analysis, we tested this question by teasing out the effects of sequential implementation of interventions. We also added the cautionary note where results were only available from fewer countries.

- On Page 21, Lines 2-6, the authors state: “The findings may also help decide which intervention to lift first as the epidemic curve starts to flatten”. I would be cautious in suggesting this - it may be too difficult to predict which intervention to lift first based on data from implementing physical interventions at the beginning of the pandemic as there are too many factors which may confound this (for instance differences in SARS-CoV-2 transmission rates, differences in other public health policies e.g. wearing of masks).

Authors’ response: **DONE.** Thank you for the comment. We agree with this, and revised the sentence accordingly.

Minor comment

- Page 12, line 53: (pooled IRR: 0.91, 95% CI: 0.82, 1.0; N=8). Confidence interval should be written with a hyphen rather than a comma.

Authors' response: **DONE**. Thank you. In line with the BMJ requirement, we reported all the CIs as “LCI to UCI”.

- Page 16, lines 31-35: “Similar study from Hubei and Guangdong also reported significant reduction in COVID-19 incidence.” Grammatical error – could be re-written as “A similar study from Hubei and Guangdong also reported a significant reduction in COVID-19 incidence”.

Authors' response: **DONE**. Thank you. It was revised as you suggested.