naturereseach

Peer Review Information

Journal: Nature Human Behaviour
Manuscript Title: Socioeconomic impacts of COVID-19 in low-income countries
Corresponding author name(s): Jeffrey D. Michler

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:

19th November 2020

Dear Dr. Michler,

Thank you once again for your manuscript, entitled "Socioeconomic impacts of COVID-19 in low-income countries", and for your patience during the peer review process.

Your Article has now been evaluated by 3 referees. You will see from their comments copied below that, although they find your work of potential interest, they have raised quite substantial concerns. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version if you are willing and able to fully address reviewer and editorial concerns.

We hope you will find the referees' comments useful as you decide how to proceed. If you wish to submit a substantially revised manuscript, please bear in mind that we will be reluctant to approach the referees again in the absence of major revisions. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

In particular, your revision must address the following (as well as all other reviewer comments):

1) All three reviewers raise the concern that your manuscript makes causal claims in the absence of a causal identification design. Your revision must be clear about the correlational nature of the data and remove causal claims.

2) Reviewers 1 and 3 both highlight the large volume of results and raise concerns about the clarity and level of detail with which they are presented. As you revise, we ask that you address these points by ensuring that all core methodological information is provided with in the paper and Supplementary Information and that only essential figures and tables are included.

3) Reviewer 3 also raises questions about the appropriateness of the modeling approach, choice of standard errors, and variables included. We ask that you address each of these concerns. With respect to your analytical strategy, our view is that simpler analyses are acceptable, but regardless of analysis, you must ensure that you appropriately correct for multiple comparisons.

4) Please ensure that the COVID-19 policy context is sufficiently described for each country, as highlighted by Reviewer 1.

In addition to these reviewer concerns, we have noted that a number of the reported results do not reach conventional levels of statistical significance, without any evidence that you a priori set your alpha level to a value other than .05. Please remove claims of significance for any results with p values higher than .05. It is journal policy that these results must be reported as non-significant.

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. I have also attached a template manuscript file that exemplifies our policies and formatting requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

If you wish to submit a suitably revised manuscript we would hope to receive it within 3 months. We understand that the COVID-19 pandemic is causing significant disruptions which may prevent you from carrying out the additional work required for resubmission of your manuscript within this timeframe. If you are unable to submit your revised manuscript within 6 months, please let us know. We will be happy to extend the submission date to enable you to complete your work on the revision.

With your revision, please:

• Include a “Response to the editors and reviewers” document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.

• Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

Thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further.

Sincerely,
Aisha

Aisha Bradshaw
Editor
Nature Human Behaviour

Reviewer expertise:

Reviewer #1: survey methods

Reviewer #2: health and development economics

Reviewer #3: development and agricultural economics

REVIEWER COMMENTS:

Reviewer #1:
Remarks to the Author:
This paper uses repeated cross-sectional survey data in four low-income African countries to estimate the impacts of the covid-19 pandemic on a range of social, psychological, and economic outcomes. These are phone surveys with samples drawn randomly from a frame of respondents to earlier face-to-face surveys. The methodology and analysis is generally well described and reported and the results are timely, interesting, and policy relevant.

The authors are extremely ambitious about the amount of information they are trying to include in a standard length journal article and it is a difficult job at times to follow what is going on. It is necessary to read external documents in various places and there a very large number of tables and figures in the paper itself. Even with all this, there are a number of places (see below) where I felt that I needed more information to be able to properly assess the findings that are presented. My feeling is that it would be better not to try to provide an exhaustive account of this many outcomes but rather to focus on a smaller selection of key variables, probably those relating to economic impacts.

More detail in the paper itself is needed on the survey designs, readers should not be directed to external documents for this key information. The methods section should describe the sampling designs of the initial face-to-face surveys that serve as the frames for the phone surveys provide their sample sizes, response rates and so on. Additional information should also be provided about the response rates and sample sizes of the phone surveys at each wave. Information about the weighting and other complex design features should also be included.

The paper needs a more detailed description of the actual restrictions that were in place in each country at the start of the paper, perhaps in a table. Without this it is difficult to interpret the survey findings. For example the paper says that Malawi did not implement stay at home orders but did close schools. What then does it mean to show in Figure A that about 50% of people knew that the government had restricted social gatherings (had it?) and that around 20% knew the government had closed schools?

The estimates of loss of income include incidences of lost income that may not have been due to covid-19. Greater clarity on how these estimates have been calculated is necessary.

It is not clear whether the estimates of food insecurity are estimates of levels or change, this should be clarified. If they are estimates of levels, how can anything meaningful be said about the impact of the pandemic on the distribution of food insecurity by income quintile? This applies to any other parts of the paper that only have post-pandemic measures – greater clarity about what can and cannot be attributed to covid is needed.

It is not clear how the estimates of ‘virus related shock’ in table s21 have been calculated. The text on page 10 refers to this as a income shock but table s21 does not reference income.

The authors make several statements about the policy relevance of their findings and I agree with them. It would therefore be useful to have at least some brief consideration of what these might be in the discussion.

Reviewer #2:
Remarks to the Author:
Socioeconomic impacts of COVID-19 in low-income countries
This paper presents the first large-scale evidence on the socio-economic impact of the COVID-19 in low economic settings. The paper uses large scale almost-nationally representative data from four African countries and covers a wide range of areas interest to policy makers in terms of the socioeconomic impact of COVID-19 – income loss, coping strategies, documenting heterogeneity across a wide range of characteristics. The paper also presents useful information on awareness of the pandemic and behaviour change that provide useful insights for understanding spread of the disease.

Overall, I think this paper makes important and timely contributions to inform our understanding of the socioeconomic impact of the pandemic. The paper is very well-written and the authors have been very transparent about the methods.
I have only minor comments:
The discussion in lines 564 to 574 seem to imply that some of the estimates could be interpreted causally. This is debatable and probably not necessary. In my view, this paper makes an important contribution to literature without needing to produce causal estimates. I will advise the authors to revise those sections to remove the implications of causality.

I find the discussions in lines 239 to 244 rather confusing. It is not clear why the authors regard the findings that “food insecurity is significantly higher in households whose school-aged children are not engaged” to be good news. That section needs some clarification.

A few typos to address:
On line 7, it appears the word “scare” should be “scarce”.
On line 161, the word “of” is missing from ‘prevalence of food”

Reviewer #3:
Remarks to the Author:
This paper provides statistics relating to the effects of COVID-19 and the associated lockdowns on households in four African countries: Ethiopia, Malawi, Nigeria and Uganda. The paper covers knowledge and awareness, COVID prevention measures, economic coping strategies, food security, access to goods and disruptions to education. The paper is mostly descriptive in nature, which I do not view as a problem. The authors do imply causality in some cases they should not. I do not think the authors appropriately describe, and in some cases use, their methodology. It also appears to me they are missing some opportunities to use pre-COVID data to make comparisons. I offer more detail below. I am also concerned the paper is light on behavioral insights. Perhaps this could be addressed by paying more attention to coping mechanisms. The paper does include a fair amount of interesting analysis on attitudes towards COVID and prevention methods but does not go into nearly enough depth regarding coping behavior in my opinion, especially for this journal. For instance, the authors could engage in a more robust discussion on the long-term impacts of various coping mechanisms or conduct some analysis to see which kinds of households resorted to different coping mechanisms.

Attribution of outcomes to COVID:
The authors often present results as causal impacts of COVID. For some outcomes, this is fair. For instance, when the outcome variable is “households reporting a decrease in income” (Figure 2A), it can safely be assumed that the decrease in income is from COVID and the associated lockdown (if the question was asked in this manner). The same goes for “change in business revenue” (Figure 2B), “concern that family or self will fall ill with COVID” (Figure 2D), and coping strategies (Figure 3A), provided the survey asked about coping strategies specific to COVID and the associated lockdown.

Other outcomes cannot really be attributed to COVID. Take the case of prevalence of food insecurity (Figure 2C). There were certainly households in the sample that were food insecure before COVID. Without a counterfactual (e.g., pre-COVID data) we cannot attribute the outcome to COVID. While there are certainly issues with making pre-post comparisons, given the nature of the COVID shock it is probably the best one can do. The pre-COVID data exists, so why not use it here? A similar argument can be made for why these other outcomes cannot be attributed to COVID without having some counterfactual against which to compare: “prevalence of household’s inability to buy medicine” (Figure 3B), “households with children engaged in learning activities” (Figure 3C), and “households with children experiencing educational contact” (Figure 3D). For these outcomes, pre-COVID data from the same households could be used to better capture the impacts of COVID rather than the current state of these variables.

Methodological issues
I do not think “sampling techniques” should be considered a method used to “empirically evaluate the effects of the pandemic on households”. Clearly sample weighting is needed, but it is not a technique to evaluate impact on its own.

I also question why the authors reference “reduced-form econometric methods”. I assume this is to make a distinction from “structural econometric methods”. It is true that the methods used here are reduced-form, but that would be completely obvious to any reader that knows what reduced-form and structural econometrics are, and completely irrelevant to someone who does not know the difference. To me, it seems the authors are just throwing in unnecessary jargon to make the statistical methods seems more sophisticated than they really are.

Taking a closer look at the econometrics the authors use and how they interpret the results, it seems they do not need econometrics at all. Comparisons between means would suffice to get most points across. Most regressions include only country fixed effects (Tables S1, S2, S3, S6, S10, S15, S23). In other cases, regressions only include survey round fixed effects (Tables S4, S5). This boils down to a comparison of means and could be presented as such. The one advantage to using a regression, I suppose, is the ability to easily cluster standard errors. I note that if the authors are arguing differences in outcomes are due to differences in country or survey round, they should cluster at the level of the fixed effect and not use Huber-White robust standard errors (Abadie et al. 2018).

Some regressions include one other dummy variable plus country fixed effects (Tables S11, S12), survey round and country fixed effects (S13), consumption quintile and country fixed effects (S27, S29), or educational activity and food insecurity (S30, which is a very strange regression to run). Why? Is country correlated with the dummy variable of interest in these cases?

The bottom line here is that I think the authors could do most of the meaningful comparisons with simple t-tests, considering the explanatory variables of interest are all binary. In some cases, it may be important to do multivariate regressions and/or use interaction terms, in which case econometric analysis makes sense. In these cases, the authors need to provide a better motivation for doing this more complex analysis, because as the paper is now, I do not think these analyses bring much insight, e.g. gender of household head x survey round in S16 and S25, or using school activity as an explanatory variable for food security (S30).

I also want to raise the issue of using female headship as an explanatory variable. What is the authors’ motivation for doing this? Female headship is correlated with many variables that the authors do not control for in any way. So, while it can be a predictive variable, it is not safe at all to attribute any differences in outcomes to the gender of the household head, all else equal. Given this, and given that the authors do not propose any policy remedies conditional on gender of the household head, I do not see a point in doing this analysis. If the authors do want to include it they need to caveat the findings appropriately.

Some other issues:
Page 25: “Figures S2 and S2…”

It is unclear why some results were selected to be put in figures in the main text whereas others were relegated to Supplementary Materials. I think the authors should be far more judicious about what tables and figures to put in the paper at all, and about which ones to put in the main text. These should be the ones with the key take-aways.
Figure S2 has food security on the horizontal axis, but this is not mentioned at all in the paper when discussing Figure S2 (page 9). I think S2 can be removed.

Author Rebuttal to Initial comments

Response to Editor

In what follows, we reproduce (in *italics*) and respond to each of your comments.

1) *All three reviewers raise the concern that your manuscript makes causal claims in the absence of a causal identification design. Your revision must be clear about the correlational nature of the data and remove causal claims.*

   We have removed causal claims from the manuscript. We are now clear that the results we present are correlational in nature.

2) *Reviewers 1 and 3 both highlight the large volume of results and raise concerns about the clarity and level of detail with which they are presented. As you revise, we ask that you address these points by ensuring that all core methodological information is provided with in the paper and Supplementary Information and that only essential figures and tables are included.*

   We have endeavored to reduce the results while at the same time expanding our discussion of the results that remain. We follow two Reviewer suggestions specifically: per Reviewer 3, we have removed all results related to gender and per Reviewer 1, we now spend more time discussing income loss, coping strategies, and food insecurity.

3) *Reviewer 3 also raises questions about the appropriateness of the modeling approach, choice of standard errors, and variables included. We ask that you address each of these concerns. With respect to your analytical strategy, our view is that simpler analyses are acceptable, but regardless of analysis, you must ensure that you appropriately correct for multiple comparisons.*

   We have worked to address Reviewer 3's concerns on these points. First, we have simplified the discussion of the modeling approach to remove jargon and make our analytical strategy more concise and easier to understand for a non-economic audience. Second, we do not believe that clustering standard errors (as opposed to estimating robust standard errors) is required. We have provided our reasoning, supported by citations, in our detailed response to Reviewer 3.

4) *Please ensure that the COVID-19 policy context is sufficiently described for each country, as highlighted by Reviewer 1.*

We have added additional information about COVID-19 policies in each country. We believe this information enriches the paper and provides greater context for interpreting results that differ across countries.

In addition to these reviewer concerns, we have noted that a number of the reported results do not reach conventional levels of statistical significance, without any evidence that you a priori set your alpha level to a value other than .05. Please remove claims of significance for any results with p values higher than .05. It is journal policy that these results must be reported as non-significant.

We have updated the text and all tables to reflect the journal policy that only $p < 0.05$ is considered significant. We now refer to anything below this critical threshold as non-significant.

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. I have also attached a template manuscript file that exemplifies our policies and formatting requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

We have worked to bring the manuscript into house style, including reformatting text, figures, tables, and references. We have made these formatting changes without tracking changes so that the more substantial changes to content are clear to the reader.

Response to Review #1

In what follows, we reproduce (in *italics*) and respond to each of your comments.

The authors are extremely ambitious about the amount of information they are trying to include in a standard length journal article and it is a difficult job at times to follow what is going on. It is necessary to read external documents in various places and there a very large number of tables and figures in the paper itself. Even with all this, there are a number of places (see below) where I felt that I needed more information to be able to properly assess the findings that are presented. My feeling is that it would be better not to try to provide an exhaustive account of this many outcomes but rather to focus on a smaller selection of key variables, probably those relating to economic impacts.

We appreciate the Reviewer's concern about the length of the paper and the amount of results presented. We have revised the paper with this in mind and have focused the updated version on issues related to income, coping, and food security. While we have not eliminated any of the main figures or findings, we have removed all the supplementary figures along with six of the 31 tables.

More detail in the paper itself is needed on the survey designs, readers should not be directed to external documents for this key information. The methods section should describe the sampling designs of the initial face-to-face surveys that serve as the frames for the phone surveys provide their sample sizes, response rates and so on. Additional information should also be provided about the response rates and sample sizes of the phone surveys at each wave. Information about the weighting and other complex design features should also be included.

We have provided substantially more details on the survey design, sampling frame, and the connection between pre- and post-COVID sample size, attrition, etc. These include Extended Data Table 1 and Extended Data Figure 2, which provide information on sampling and response rate for all survey rounds. This information is provided in the Methods section, as part of the article. Due to space constraints, we also include more information about original sampling frames, sampling methods, etc. in the Supplementary Information.

The paper needs a more detailed description of the actual restrictions that were in place in each country at the start of the paper, perhaps in a table. Without this it is difficult to interpret the survey findings. For example, the paper says that Malawi did not implement stay at home orders but did close schools. What then does it mean to show in Figure A that about 50% of people knew that the government had restricted social gatherings (had it?) and that around 20% knew the government had closed schools?

We have worked to add additional information regarding the actual restrictions that were in place in each country. Per the Reviewer's specific example, the Government of Malawi tried to issue and enforce a 21-day lockdown, ordering all but essential workers to stay in their houses. This order was successfully challenged in the High Court. The result was that the Government of Malawi could only issue recommendations and advise citizens to stay at home and avoid social gatherings. The one thing they could enforce was closures of public schools. The end result of all this was a degree of confusion and uncertainty about what the Government had instituted. 50% of people knew that the government had advised against social gatherings (the question is "advised to avoid gatherings") while only 20% were aware that the government had closed schools. This low knowledge of school closures is in part due to the fact that this is relevant only for households with school age children. Across all four countries, knowledge of school closures tends to be lower than knowledge of other government actions. We have worked to add these and other additional details to the paper.

The estimates of loss of income include incidences of lost income that may not have been due to covid-19. Greater clarity on how these estimates have been calculated is necessary.

The questions about income are phrased in such a way as to provide insight on income loss since the pandemic. Households are first asked "In the last 12 months, which of the following were your household's sources of livelihood?" If a respondent answers "Yes" to any of the sources listed, they are then asked the follow-up question about that source: "Since mid-March has income from [SOURCE] increased/decreased/stayed the same?" While it is possible that income loss is due to non-COVID-19 related events, as Reviewer 3 notes (and we agree): tying the loss to the mid-March time period, when countries began to impose restrictions provides a fair degree of certainty that income loss is a result of the pandemic. We have revised the section on income to provide greater clarity on the phrasing of the questions.

It is not clear whether the estimates of food insecurity are estimates of levels or change, this should be clarified. If they are estimates of levels, how can anything meaningful be said about the impact of the pandemic on the distribution of food insecurity by income quintile? This applies to any other parts of the paper that only have post-pandemic measures – greater clarity about what can and cannot be attributed to COVID-19 is needed.

We have worked to clarify and elaborated on the food insecurity estimates. The estimates presented are in levels, not in changes. Where possible, we have added more information about the changes in food insecurity before and since the pandemic. This is now included as Extended Data Figure 2, which presents food insecurity in Nigeria before and since the pandemic. We believe that there is value in knowing the status of food insecurity during the pandemic, even if we are not drawing inference about the change in food insecurity. It is of interest to governments, stakeholders, and other researchers to understand the current status of food insecurity in these countries.

It is not clear how the estimates of 'virus related shock' in table s21 have been calculated. The text on page 10 refers to this as an income shock but table s21 does not reference income.

Thank you for catching this inconsistency. In the paper and table, we interchangeably referred to "virus related shocks" and "income shocks." This was because the surveys ask about shocks impacting a household's income earning potential since mid-March. We realize this inconsistency in wording caused confusion among readers. We have added information to the paper in order to clarify the types of shocks asked about and that these shocks occur after the start of the pandemic. We no longer refer to "income shocks" or "virus-related shocks" but simply "shocks to the household." We trust this clarifies any confusion.

The authors make several statements about the policy relevance of their findings and I agree with them. It would therefore be useful to have at least some brief consideration of what these might be in the discussion.

We have added a richer discussion of the policy relevance of our findings to the paper. This appears in the Discussion of the manuscript and includes discussion of the opportunities for strengthened social protections and improved technology, sustained citizen engagement, and ongoing communication and messaging from governments, in order to mitigate and address the negative impacts of the COVID-19 pandemic in these countries. These include the opportunity to strengthen social protections to address income losses, food insecurity, and difficulties with accessing basic necessities, as well as the potential benefits to increased access to technology, specifically to support remote education efforts. We further identify the ongoing need of governments to communicate clearly, dispelling COVID-19 myths, and encouraging safe practices to mitigate the transmission of the disease.

Response to Review #2

In what follows, we reproduce (in *italics*) and respond to each of your comments.

The discussion in lines 564 to 574 seem to imply that some of the estimates could be interpreted causally. This is debatable and probably not necessary. In my view, this paper makes an important contribution to literature without needing to produce causal estimates. I will advise the authors to revise those sections to remove the implications of causality.

We appreciate the Reviewer's concern that we have oversold our ability to identify causal impacts. Per your comments, and similar comments from the other two Reviewers, we have revised our discussion of the econometric model and removed language that implies causality.

I find the discussions in lines 239 to 244 rather confusing. It is not clear why the authors regard the findings that "food insecurity is significantly higher in households whose school-aged children are not engaged" to be good news. That section needs some clarification.

Per comments from other Reviewers, we have removed this discussion from the paper.

A few typos to address:

On line 7, it appears the word "scare" should be "scarce".

On line 161, the word "of" is missing from 'prevalence of food"

Thanks for noting these. We have addressed these issues in the new draft.

Response to Review #3

In what follows, we reproduce (in *italics*) and respond to each of your comments.

Attribution of outcomes to COVID:

The authors often present results as causal impacts of COVID. For some outcomes, this is fair. For instance, when the outcome variable is "households reporting a decrease in income" (Figure 2A), it can safely be assumed that the decrease in income is from COVID and the associated lockdown (if the question was asked in this manner). The same goes for "change in business revenue" (Figure 2B), "concern that family or self will fall ill with COVID" (Figure 2D), and coping strategies (Figure 3A), provided the survey asked about coping strategies specific to COVID and the associated lockdown. Other outcomes cannot really be attributed to COVID. Take the case of prevalence of food insecurity (Figure 2C). There were certainly households in the sample that were food insecure before COVID. Without a counterfactual (e.g., pre-COVID data) we cannot attribute the outcome to COVID. While there are certainly issues with making pre-post comparisons, given the nature of the COVID shock it is probably the best one can do. The pre-COVID data exists, so why not use it here? A similar argument can be made for why these other outcomes cannot be attributed to COVID without having some counterfactual against which to compare: "prevalence of household's inability to buy medicine" (Figure 3B), "households with children engaged in learning activities" (Figure 3C), and "households with children experiencing educational contact" (Figure 3D). For these outcomes, pre-COVID data from the same households could be used to better capture the impacts of COVID rather than the current state of these variables.

We appreciate the Reviewer's concern that we have done a poor job in clarifying what outcomes can be attributed to COVID and what outcomes cannot. This was a concern raised by the other two Reviewers and we have worked hard to address the concern by clarifying where we can, with some confidence, assume outcomes are the result of COVID and the associated lockdowns.

As the Reviewer points out, the cause-effect relationship is straightforward for outcomes in Figure 2A (income loss), 2B (business revenue loss), 2D (concern), and 3A (coping strategies). In all four cases the questions asked about changes to the household's situation since mid-March. We also believe that Figure 3C (educational contact) falls into this category. The surveys asked, "Were any children attending school before schools were closed due to coronavirus?" If a respondent answered "yes" to this question, then they were asked "Have the children been engaged in any education or learning activities in the last week?" Therefore, the measure of engagement is contingent on children in the household having been in school prior to the lockdowns.

For Figure 3D (educational contact), the causal chain is less clear, though that we have a temporal dimension to this question, and that we see change over time. This suggests to us that levels of contact were higher prior to COVID. In order to be on the cautious side, we have revised this paragraph in the manuscript so that it only reports levels of contact and not changes from pre-COVID levels.

For the remaining figures, 2C (food insecurity) and 3B (access), we agree with the Reviewer that the questions are worded in such a way that outcomes cannot be attributed to COVID. We have revised the discussion of each figure to clarify this. In addition to clarifying that these outcomes are not caused by COVID, we have worked to enrich the discussion and add additional evidence on food security and access to these necessary goods.

For food insecurity, pre-COVID data exists for households in Nigeria (the pre-COVID LSMS-ISA surveys in the other countries did not use FIES to measure food insecurity). Using this Nigeria data, we have also added an "Extended Data Figure 2", which presents food insecurity in Nigeria before and since the pandemic. This allows us to make inference using household fixed effects on the COVID-related change in Nigeria, but not in any of the other countries. Despite this limitation, we feel that there is value in knowing the status of food insecurity during the pandemic, even if we are not drawing inference about the change in food insecurity due to the pandemic. For access, we provide more information on the wording of the questions and household responses to why they were unable to access food, medicine, and/or soap.

Methodological issues

I do not think "sampling techniques" should be considered a method used to "empirically evaluate the effects of the pandemic on households". Clearly sample weighting is needed, but it is not a technique to evaluate impact on its own.

In revising the paper, we have removed this statement.

I also question why the authors reference "reduced-form econometric methods". I assume this is to make a distinction from "structural econometric methods". It is true that the methods used here are reduced-form, but that would be completely obvious to any reader that knows what reduced-form and structural econometrics are, and completely irrelevant to someone who does not know the difference. To me, it seems the authors are just throwing in unnecessary jargon to make the statistical methods seems more sophisticated than they really are.

Yes, the Reviewer is correct, the reference to reduced-form econometrics was to distinguish it from structural methods. In reading other economic papers in Nature and Nature Human Behavior, this appeared as common phrasing, but we agree it is jargon and not necessary. We have replaced "reduce-form econometric methods" with "simple econometric methods."

Taking a closer look at the econometrics the authors use and how they interpret the results, it seems they do not need econometrics at all. Comparisons between means would suffice to get most points across. Most regressions include only country fixed effects (Tables S1, S2, S3, S6, S10, S15, S23). In other cases, regressions only include survey round fixed effects (Tables S4, S5). This boils down to a comparison of means and could be presented as such. The one advantage to using a regression, I suppose, is the ability to easily cluster standard errors. I note that if the authors are arguing differences in outcomes are due to differences in country or survey round, they should cluster at the level of the fixed effect and not use Huber-White robust standard errors (Abadie et al. 2018).

We appreciate the Reviewer's concern about cluster robust versus robust standard errors. While we agree that it is standard practice to cluster at the level of the fixed effects (country or wave, in our case), we believe that the sampling design and the population on which we make inference does not require correction of standard errors for with-in cluster correlation – correcting for heteroskedasticity is sufficient. Our view is informed by Abadie et al. (2017, When Should You Adjust Your Standard Errors for Clustering?). Per Abadie et al. (p. 2), clustering will be necessary if "there are clusters in the population of interest that are not represented in the sample." The classic example here would be sampling from 100 villages in a country and then sampling from households in a village. In this case, if one wants to make inference about the population of the country, one would need to cluster on village, since not all villages are present in the sample. Conversely, if one has observations from all villages or if one simply wanted to make inference on the sample (not the population), clustering would not be necessary. Abadie et al. (p. 7) write, "The LZ [Liang-Zeger cluster] standard errors are based on the presumption that there are clusters in the population of interest beyond the 100 clusters that are seen in the sample. The EHW [Eicker-Huber-White] standard errors assume the sample is drawn randomly from the population of interest. It is this presumption underlying the LZ standard errors of existence of clusters that are not observed in the sample, but that are part of the population of interest, that is critical." In our case, we have observations from all four countries for which we make inference on. Because of this, we do not believe that clustering is called for and adjusting for heteroskedasticity in the data is sufficient.

While we believe that our sampling design does not call for clustering, in the interests of being as complete, thorough, and transparent as possible, we examined the effect that clustering would have on our results. In all cases that we checked, clustering by country resulted in smaller standard errors and more significant results than Huber-White robust errors. We believe the reason for this is because we have an extremely small number of clusters (four in the case of country FEs and two in the case of round FEs). Per Cameron and Miller (2015) and Cameron, Gelbach, and Miller (2017), analytical asymptotics can be off when the number of clusters is below 30. The extremely small standard errors in our results may be due to the small number of clusters. Webb (2014) and MacKinnon and Webb (2018) suggest that wild bootstrapping can be used to calculate standard errors in cases with small numbers of clusters (six or more clusters). Young (2019) suggests that randomization inference can also be used to calculate clustered standard errors when the number of clusters are small and analytical asymptotics are distorted. We implemented both of these methods to determine their effect on the small size of the clustered standard errors. We followed Roodman et al. (2019) to implement the wild bootstrap and Hess (2017) to implement randomization inference. Neither process was terribly successful. p-values that had been in the 0.05 to 0.01 range when calculated from Huber-White robust standard errors shrunk to < 0.01 when calculated from cluster robust standard errors. Using the wild bootstrap resulted in p-values > 0.20, while randomization inference resulted in p-values > 0.50. The extremely large size of these p-values, and their large divergence from analytical p-values, suggests to us that computational methods such as the wild bootstrap randomization inference are unable to overcome the extremely small number of clusters. For Cameron and Miller (2015) and Webb (2014), six clusters seems to be the lower bound for numerical methods to work correctly. Our tests with numerical methods on four and two clusters do not return p-values of reasonable size, given the magnitude of the coefficients and the number of observations in the data.

The preceding discussion about clustered standard errors should not draw away from the fact that we believe Huber-White robust standard errors are the appropriate adjustment to be made for our data and population of interest. We present the preceding discussion in the interest of completeness, transparency, and to allay any concerns the Reviewer may have that we did not take these concerns seriously.

Some regressions include one other dummy variable plus country fixed effects (Tables S11, S12), survey round and country fixed effects (S13), consumption quintile and country fixed effects (S27, S29), or educational activity and food insecurity (S30, which is a very strange regression to run). Why? Is country correlated with the dummy variable of interest in these cases?

Yes, the Reviewer is correct, we control for country in these regressions because we are concerned about correlation between country and the dummy variable. This applies to the rural/urban dummy (since each country has different degrees of urbanization), to concerns (since disease prevalence varies by country), and to round (the timing of each round differs by country). This also applied to gender of household head, though we have now dropped all of those regressions per the Reviewer's suggestion.

The reviewer is correct that this logic does not apply when we are testing for differences in consumption quintile. Consumption quintiles are calculated within each country and thus are uncorrelated with country fixed effects. We have now dropped country fixed effects from these regressions (S17, S23, and S25). As one would expect, since country fixed effects are uncorrelated with quintiles, the results do not change when we drop country fixed effects.

The bottom line here is that I think the authors could do most of the meaningful comparisons with simple t-tests, considering the explanatory variables of interest are all binary. In some cases, it may be important to do multivariate regressions and/or use interaction terms, in which case econometric analysis makes sense. In these cases, the authors need to provide a better motivation for doing this more complex analysis, because as the paper is now, I do not think these analyses bring much insight, e.g. gender of household head x survey round in S16 and S25, or using school activity as an explanatory variable for food security (S30).

Yes, the Reviewer is correct, in several cases the regression results provide statistically identical results to simple t-tests. However, in many cases, the p-values change when we calculate the Huber-White standard errors. We believe that using the multivariate regressions to test differences is superior to simple t-tests for two reasons. First, correcting for heteroskedasticity does matter in our context. Second, we can calculate the same number of tests much more efficiently using multivariate regression, as the regression allows for testing pairs simultaneously, as opposed to the sequential nature required with t-tests.

I also want to raise the issue of using female headship as an explanatory variable. What is the authors' motivation for doing this? Female headship is correlated with many variables that the authors do not control for in any way. So, while it can be a predictive variable, it is not safe at all to attribute any differences in outcomes to the gender of the household head, all else equal. Given this and given that the authors do not propose any policy remedies conditional on gender of the household head, I do not see a point in doing this analysis. If the authors do want to include it they need to caveat the findings appropriately.

We had included information about female headship based on feedback from presenting results at workshops and seminars. That said, we agree with the reviewer that female headship is likely correlated with a number of unobservables. This means, as the Reviewer points out, that we cannot attribute all of the difference in headship to gender. Because of this, we have removed all tables, figures, and discussion relating to female headship.

Some other issues:

Page 25: "Figures S2 and S2…"

We have removed Figure S2 so this is no longer an issue.

It is unclear why some results were selected to be put in figures in the main text whereas others were relegated to Supplementary Materials. I think the authors should be far more judicious about what tables and figures to put in the paper at all, and about which ones to put in the main text. These should be the ones with the key take-aways. Figure S2 has food security on the horizontal axis, but this is not mentioned at all in the paper when discussing Figure S2 (page 9). I think S2 can be removed.

Per the Reviewer's suggestion, we have removed the supplementary figures in order to clarify the key takeaways and remove the results pertaining to the gender of the head of household.

Decision Letter, first revision:

4th December 2020

Dear Dr Michler,

RE: "Socioeconomic impacts of COVID-19 in low-income countries"

Thank you for submitting your revised manuscript and for all your work on the revision.

Although your manuscript has been revised in response to reviewer comments, it does not fully comply with our editorial policies and formatting requirements. In particular, inferential statistical results are not yet fully reported according to our requirements. For all results discussed in text, we ask that authors include coefficients/effect sizes, exact p-values (unless <0.001), and confidence intervals. In cases where the text discusses a number of results from regression models (as is the case for much of your paper), we do not require full in-text reporting of each result. However, the reader should be clearly pointed to tables with the full results, and all of the required statistical information should be included in these tables tables. It is our policy that tables must provide exact p-values, rather than using asterisks to mark significance.

Before we can send the manuscript back to our reviewers, we ask that you revise it to ensure that it complies fully with our policies for reporting inferential statistical results, by revising your tables to include the required information. I have attached another copy of our checklist, as well as the template document that exemplifies our formatting and policy requirements. If you are uncertain as to how to address any of the points related to statistical reporting in the checklist, please don't hesitate to contact me.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

Thank you in advance for attending to these requests and I look forward to receiving your revised manuscript.

Sincerely,
Aisha

Aisha Bradshaw
Editor
Nature Human Behaviour

Decision Letter, second revision:

19th January 2021

Dear Jeff,

Thank you once again for your revised manuscript, entitled "Socioeconomic impacts of COVID-19 in low-income countries," and for your patience during the peer review process.

Your manuscript has now been evaluated by the same 3 reviewers who saw the original version of your study, and you will find their comments included at the end of this letter. Although the reviewers continue to find your work to be of interest, and Reviewer 2 is now fully satisfied with your study, they also raise some important ongoing concerns. We remain interested in the possibility of publishing your study in Nature Human Behaviour, but would like to consider your response to these concerns in the form of a further revision before we make a decision on publication.

You will see from their comments that Reviewers 1 and 3 highlight remaining concerns about the analyses and interpretations of results in your manuscript. Reviewer 3 in particular indicates that their previous concerns about the methodological approach and the handling of standard errors have not been addressed appropriately in this revision. In order to identify the most appropriate resolution for these issues, we submitted your manuscript to our reviewer consultation process, in which we sought Reviewer 1's feedback on Reviewer 3's outstanding concerns. Reviewer 1 noted that, although reporting comparisons of means with t-tests would result in a simpler analytical approach, your existing approach is appropriate; and that your treatment of standard errors is appropriate, as, on the assumption that the samples in the four countries are each drawn with equal probability of selection, there is no non-independence of sample units within countries. You will find Reviewer 1's full feedback included with the reviews at the end of this letter. Based on this feedback, we ask that you retain your regression analyses and treatment of standard errors as they currently stand.

Although we do not require country-clustered standard errors or a change in overall analytical approach, we do require that you thoroughly address all other reviewer points, including the need for population weighting, concerns about the strength of the food insecurity analysis, and the need to avoid causal claims and implications.

Finally, your revised manuscript must continue to comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a copy of our checklist that lists all of our requirements. I have also attached a template manuscript file that exemplifies our policies and formatting requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

In sum, we invite you to revise your manuscript taking into account all reviewer and editor comments. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

We hope to receive your revised manuscript within four to eight weeks. We understand that the COVID-19 pandemic is causing significant disruption for many of our authors and reviewers. If you cannot send your revised manuscript within this time, please let us know - we will be happy to extend the submission date to enable you to complete your work on the revision.

With your revision, please:

• Include a “Response to the editors and reviewers” document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.

• Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Sincerely,
Aisha

Aisha Bradshaw
Editor
Nature Human Behaviour

Reviewer expertise:

Reviewer #1: survey methods

Reviewer #2: health and development economics

Reviewer #3: development and agricultural economics

REVIEWER COMMENTS:

Reviewer #1:
Remarks to the Author:
This is my review of the revised version of the manuscript, “Socioeconomic impacts of COVID-19 in low-income countries”. The authors have done a good job of addressing the concerns raised by myself and the other two reviewers, particularly in making the paper more focused and scaling back their ambitions for what can be covered in a journal article. Nonetheless, there remain some problems that need to be addressed before the paper can be published in Nature Human Behavior.

The main statistical issue relates to the estimates of quantities across the 4 countries, which are an important part of the paper (to take one example, the 61% of adult population suffering moderate or severe food insecurity L173). It is not clear whether these estimates have been weighted by country population size, I could see no reference to this in the paper. If the authors have not applied such a weight then these estimates will over-represent the populations in the smaller countries like Malawi and under-represent the populations in the larger countries like Nigeria. The sample sizes of the surveys have not been drawn proportionately to the sizes of the population in each country so this needs to be done via weighting at the analysis stage. Consider a global survey of all countries where the sample size in each country is approximately 1000. If a global average is produced by taking the unweighted arithmetic mean across all countries then an individual in China will be given equal weight to an individual in Iceland, despite the vast difference in the sizes of the respective populations.

Second, while the revised manuscript is now more careful about claims of causal impact of covid 19 on the survey variables, there are still problems with how this is treated. Sometimes this is simply the choice of language, for example L105 refers to ‘impacts’ of covid. In other places it is more fundamental. For example, there is a problem with assuming that all lost income since March is due to the pandemic. At the very least this must be considered an upper bound. That is to say that some households would have lost income from some sources between March and the time of the survey in the absence of the pandemic so we cannot attribute all lost income since March to coronavirus.

Elsewhere, breaking down food insecurity estimates by pre-covid consumption quintiles is presented as a mitigation of the problem of not having pre-covid estimates of food insecurity. It is not and should not be presented as such. The authors are correct that levels of food insecurity are of interest in and of themselves but they are not able to produce estimates of change in food insecurity for three of the four countries and presenting the estimates by consumption quintiles does not change this.

As per the comment above, I am unconvinced that the estimates presented from L230 can be attributed solely to the pandemic. Presumably there would be a non-zero % of people in these countries who cannot access medicines, purchase staples and so on in pre-covid times. Greater care is therefore needed in how these estimates are presented and interpreted.

I am not persuaded by the claim in the discussion L208 that the analysis provides insight on the effects of government policies. Where is any link made in the analysis between variation in government policy between countries and the estimates presented? The only place I can see that this is done is where it is argued that awareness of government actions and precautions against infection in Malawi is lower because the government's lockdown policy was overturned in the courts. But this linkage between policy and perception is, at best, weak. Stronger evidence of a link between the estimates presented and variation in government policy is needed or this claim should be removed.

The authors should be more careful about referring to lockdown restrictions as causing economic problems (e.g. 'government efforts to limit the spread of SARS are having profound economic impacts' l19). It is clear that the immediate consequences of lockdowns are economically damaging but the counterfactual here is not normal life and there are good reasons to think that the economic consequences of taking no action would be considerably worse.

Some explanation of why these four countries were selected would be helpful, even if it was just because a sampling frame in the form of pre-covid surveys was only available for these countries.

L38 'household individual and child levels'. Households and individuals are different levels of analysis but children also constitute an individual level element.

L65 "we use well established sampling techniques" should presumably refer to weighting adjustments not sampling techniques (this phrase is repeated in the supplementary appendix).

L80 what does 'near real-time' mean?

L125 'lack of' should be 'lower'

Chart D should be presented using single bars indicating % correct rather than separate bars for true and false.

As noted above, no reason is suggested for the lower levels of knowledge/awareness in Malawi but the section heading suggests that it is due to 'government action'. This heading would be more accuarte as 'awareness and knowledge of covid-19' or similar.

Figure 2(d) is unclear. Presumably the %s of yes and no responses sum to 100% so why show bars for yes and no? Only a bar for yes is needed. 'Moderate or severe' and 'moderate' – are these groups independent or can the same respondent be in both groups?

The reference to student teacher contact is ambiguous – what constitutes contact here and over what time frame?

In the discussion I do not consider the fact that there is a lag between data collection and reporting of survey findings to be a particular limitation, nor is it a limitation that these data do not provide a complete account of the pandemic in these countries. These data provide a snapshot of what was happening at this particular point in time, that is not a limitation. It would be better to use this space to discuss the more significant limitations relating to the cross-sectional nature of much of the evidence and the difficulty of linking the estimates presented to the impact of the pandemic.

I found Extended data table 1 giving the sample sizes and response rates of the surveys in each country to be unclear. Some explanation is needed for why the number of attempted interviews is less than the number of households available on the frame and why this varies across rounds. Additionally, I assume that the response rates to the pre-covid surveys were less than 100% so the net response rates for each round should account for the initial nonresponse.

Additional comments from Reviewer 1 during consultation process:

I have read the reviewer response to the author's response to the first review.

On the matter of whether regression models or comparison of means with t-tests should be used, this is really just a matter of taste. The reviewer prefers a 'simpler' analytical approach on the basis, presumably, that it will be easier for less technically oriented readers to follow. On the other hand, the regression approach is more flexible for estimating standard errors and including covariates and regression can hardly be considered an unusual or 'difficult' method. I have some sympathy with the idea of aiming for maximum analytic parsimony but I think the reviewer is rather over-stepping their remit in asking the authors to write a paper the reviewer thinks they should have written, rather than the one they chose to write. I would therefore not support requiring them to re-write the paper using t-tests, not least as this will slow up the publication of what is a timely and important paper, without altering the actual findings or conclusions.

On the matter of the standard errors, I do not agree with reviewer 2 that these should be clustered at the country level. To understand why this is so, we need to step back and ask why this procedure is necessary in the first place. It is necessary when a sample design does not give equal selection probabilities to all population units by drawing the sample in stages, otherwise known as a cluster design. Because people tend to be more similar to one another within clusters (e.g. local areas, schools) than they are in the population, treating such data as if they are a simple random sample will tend to under-estimate the true population variance. There are various adjustments that are available for this purpose including the approach that the authors use of robust standard errors which is generally favoured by economists, while statisticians tend to prefer hierarchical models. The end result though should be the same of somewhat increasing the standard errors compared to an approach which wrongly treats the sample design as srs. So, in the case where we have 4 countries and for simplicity assume the samples are each drawn with equal probability of selection, there is no non-independence of sample units within countries. One can of course specify a country level random effect and this is indeed very common, particularly when the analyst is interested in country level covariates and for this a random effect is necessary. However, without country level variables in the model this is not necessary for obtaining correct standard errors on the level 1 fixed effects.

Reviewer #2:
Remarks to the Author:
The authors have satisfactorily addressed all my comments. I do not have additional concerns.

Reviewer #3:
Remarks to the Author:
Please see attached.

Unfortunately, the paper has not improved much since the first submission. It still uses unnecessarily complex and incorrectly applied econometric methods when t-tests would be sufficient to get most of the points across. I cannot tell if the results in the main figures (those not in the supplementary material) are even using those regression results, or if they are just presenting descriptive statistics (I see no confidence intervals here). If the econometrics are only being used for results presented in the supplementary material, why use them at all?

I like this paper as a timely snapshot of how COVID is impacting people in these four African countries. I do not like this paper as an econometric analysis of differences in impacts across countries. The econometrics is useful for some of the analysis described in the paper (specifically, differences across consumption quartiles). But other than that, I really do not see the need to use it and I think the description of the econometric methodology takes up far too much valuable space in what is a very short but dense paper.

The organization of this paper is very strange. I realize papers for *Nature Human Behavior* and other such journals differs from the structure of papers in economics journals. The structure of this paper seems to be a mix of the two. For instance, the "Results" section starts off with two paragraphs about sampling and two paragraphs about data collection. These are not results; they belong with the methods.

I have more comments below (in bold) that are either replies to the authors' rebuttal or follow-ups on my previous comments.

Response to Review #3

In what follows, we reproduce (in *italics*) and respond to each of your comments. ***Attribution of outcomes to COVID:***

The authors often present results as causal impacts of COVID. For some outcomes, this is fair. For instance, when the outcome variable is "households reporting a decrease in income" (Figure 2A), it can safely be assumed that the decrease in income is from COVID and the associated lockdown (if the question was asked in this manner). The same goes for "change in business revenue" (Figure 2B), "concern that family or self will fall ill with COVID" (Figure 2D), and coping strategies (Figure 3A), provided the survey asked about coping strategies specific to COVID and the associated lockdown. Other outcomes cannot really be attributed to COVID. Take the case of prevalence of food insecurity (Figure 2C). There were certainly households in the sample that were food insecure before COVID. Without a counterfactual (e.g., pre-COVID data) we cannot attribute the outcome to COVID. While there are certainly issues with making pre-post comparisons, given the nature of the COVID shock it is probably the best one can do. The pre-COVID data exists, so why not use it here? A similar argument can be made for why these other outcomes cannot be attributed to COVID without having some counterfactual against which to compare: "prevalence of household's inability to buy medicine" (Figure 3B), "households with children engaged in learning activities" (Figure 3C), and "households with children experiencing educational contact" (Figure 3D). For these outcomes, pre-COVID data from the same *households could be used to better capture the impacts of COVID rather than the current state of these variables.*

We appreciate the Reviewer's concern that we have done a poor job in clarifying what outcomes can be attributed to COVID and what outcomes cannot. This was a concern raised by the other two Reviewers and we have worked hard to address the concern by clarifying where we can, with some confidence, assume outcomes are the result of COVID and the associated lockdowns.

As the Reviewer points out, the cause-effect relationship is straightforward for outcomes in Figure 2A (income loss), 2B (business revenue loss), 2D (concern), and 3A (coping strategies). In all four cases the questions asked about changes to the household's situation since mid-March. We also believe that Figure 3C (educational contact) falls into this category. The surveys asked, "Were any children attending school before schools were closed due to coronavirus?" If a respondent answered "yes" to this question, then they were asked "Have the children been engaged in any education or learning activities in the last week?" Therefore, the measure of engagement is contingent on children in the household having been in school prior to the lockdowns.

For Figure 3D (educational contact), the causal chain is less clear, though that we have a temporal dimension to this question, and that we see change over time. This suggests to us that levels of contact were higher prior to COVID.

In order to be on the cautious side, we have revised this paragraph in the manuscript so that it only reports levels of contact and not changes from pre-COVID levels.

For the remaining figures, 2C (food insecurity) and 3B (access), we agree with the Reviewer that the questions are worded in such a way that outcomes cannot be attributed to COVID. We have revised the discussion of each figure to clarify this. In addition to clarifying that these outcomes are not caused by COVID, we have worked to enrich the discussion and add additional evidence on food security and access to these necessary goods.

For food insecurity, pre-COVID data exists for households in Nigeria (the pre-COVID LSMS-ISA surveys in the other countries did not use FIES to measure food insecurity). Using this Nigeria data, we have also added an "Extended Data Figure 2", which presents food insecurity in Nigeria before and since the pandemic. This allows us to make inference using household fixed effects on the COVID-related change in Nigeria, but not in any of the other countries. Despite this limitation, we feel that there is value in knowing the status of food insecurity during the pandemic, even if we are not drawing inference about the change in food insecurity due to the pandemic.

Given how much is already packed into this short paper, I disagree that there is value knowing the status of food insecurity during the pandemic without knowing the status of food insecurity before the pandemic. You can easily make the paragraph starting on line 170 about Nigeria only, because that is the only country for which you have appropriate data to make this analysis. Looking at food insecurity by consumption quartile offers no further insight on the role of COVID: under any circumstances you would expect more food insecurity among those with lower consumption. Heterogeneity in food security across countries is meaningless for this analysis. The evidence on Nigeria is striking, and throwing in this other stuff dilutes this part of the paper.

For access, we provide more information on the wording of the questions and household responses to why they were unable to access food, medicine, and/or soap.

Ok. You can do this without looking at the food insecurity numbers for the countries other than Nigeria.

Methodological issues

I do not think "sampling techniques" should be considered a method used to "empirically evaluate the effects of the pandemic on households". Clearly sample weighting is needed, but it is not a technique to evaluate impact on its own.

In revising the paper, we have removed this statement.

I also question why the authors reference "reduced-form econometric methods". I assume this is to make a distinction from "structural econometric methods". It is true that the methods used here are reduced-form, but that would be completely obvious to any reader that knows what reduced-form and structural econometrics are, and completely irrelevant to someone who does not know the difference. To me, it seems the authors are just throwing in unnecessary jargon to make the statistical methods seems more sophisticated than they really are.

Yes, the Reviewer is correct, the reference to reduced-form econometrics was to distinguish it from structural methods. In reading other economic papers in *Nature* and *Nature Human Behavior*, this appeared as common phrasing, but we agree it is jargon and not necessary. We have replaced "reduce-form econometric methods" with "simple econometric methods."

I suggest "We then use econometric methods common to the field…" or "We then use common econometric methods…". "Simple" is vague. Along these lines, there is no reason to present the OLS estimator in this paper. Lines 527-544 are a waste of valuable space. Anyone who knows anything about econometrics knows this estimator. For everyone else, this passage will be completely meaningless.

Taking a closer look at the econometrics the authors use and how they interpret the results, it seems they do not need econometrics at all. Comparisons between means would suffice to get most points across. Most regressions include only country fixed effects (Tables S1, S2, S3, S6, S10, S15, S23). In other cases, regressions only include survey round fixed effects (Tables S4, S5). This boils down to a comparison of means and could be presented as such.

I stand by this. The paper would be better if the econometrics were mostly if not totally removed and it was written as a descriptive paper. The econometric methodology certainly gets much more attention than it deserves in the paper given it is (a) very simple, and (b) in most cases not necessary beyond a t-test.

The one advantage to using a regression, I suppose, is the ability to easily cluster standard errors. I note that if the authors are arguing differences in outcomes are due to differences in *country or survey round, they should cluster at the level of the fixed effect and not use Huber-White robust standard errors (Abadie et al. 2018).*

We appreciate the Reviewer's concern about cluster robust versus robust standard errors. While we agree that it is standard practice to cluster at the level of the fixed effects (country or wave, in our case), we believe that the sampling design and the population on which we make inference does not require correction of standard errors for with-in cluster correlation – correcting for heteroskedasticity is sufficient. Our view is informed by Abadie et al. (2017, When Should You Adjust Your Standard Errors for Clustering?). Per Abadie et al. (p. 2), clustering will be necessary if "there are clusters in the population of interest that are not represented in the sample." The classic example here would be sampling from 100 villages in a country and then sampling from households in a village. In this case, if one wants to make inference about the population of the country, one would need to cluster on village, since not all villages are present in the sample.

This is the case with the LSMS data, is it not? If it is, then even if you make the argument that clustering should be informed by sample construction and not the level of treatment, then you need to cluster at the village level rather than use Huber-White robust standard errors.

Conversely, if one has observations from all villages or if one simply wanted to make inference on the sample (not the population), clustering would not be necessary. Abadie et al. (p. 7) write, "The LZ [Liang-Zeger cluster] standard errors are based on the presumption that there are clusters in the population of interest beyond the 100 clusters that are seen in the sample.

The EHW [Eicker-Huber-White] standard errors assume the sample is drawn randomly from the population of interest. It is this presumption underlying the LZ standard errors of existence of clusters that are not observed in the sample, but that are part of the population of interest, that is critical." In our case, we have observations from all four countries for which we make inference on. Because of this, we do not believe that clustering is called for and adjusting for heteroskedasticity in the data is sufficient.

This argument is incorrect. You do not have random samples of individuals taken from the populations of these countries. There is multi-level sampling, which means that you would need to cluster at the village level if you are determining your clustering based on sampling. But the reason you need to cluster your standard errors at the country level when you estimate country fixed effects as your outcomes of interest is that treatment is "assigned" at the country level. This point is clear on page 1 of the Abadie et al. (2017) paper: "Instead of a sampling issue, clustering can also be an *experimental design* issue, when clusters of units, rather than units, are assigned to a treatment". In this paper you make claims about outcomes being different in Malawi because Malawi had different policies about COVID. Thus, the treatment is assigned at the country level. Thus you need to cluster at the country level.

While we believe that our sampling design does not call for clustering, in the interests of being as complete, thorough, and transparent as possible, we examined the effect that clustering would have on our results. In all cases that we checked, clustering by country resulted in smaller standard errors and more significant results than Huber-White robust errors. We believe the reason for this is because we have an extremely small number of clusters (four in the case of country FEs and two in the case of round FEs).

This is probably the case.

Per Cameron and Miller (2015) and Cameron, Gelbach, and Miller (2017), analytical asymptotics can be off when the number of clusters is below 30. The extremely small standard errors in our results may be due to the small number of clusters. Webb (2014) and MacKinnon and Webb (2018) suggest that wild bootstrapping can be used to calculate standard errors in cases with small numbers of clusters (six or more clusters). Young (2019) suggests that randomization inference can also be used to calculate clustered standard errors when the number of clusters are small and analytical asymptotics are distorted. We implemented both of these methods to determine their effect on the small size of the clustered standard errors. We followed Roodman et al. (2019) to implement the wild bootstrap and Hess (2017) to implement randomization inference. Neither process was terribly successful. p-values that had been in the 0.05 to 0.01 range when calculated from Huber-White robust standard errors shrunk to < 0.01 when calculated from cluster robust standard errors. Using the wild bootstrap resulted in p-values > 0.20, while randomization inference resulted in p-values > 0.50. The extremely large size of these p-values, and their large divergence from analytical p-values, suggests to us that computational methods such as the wild bootstrap randomization inference are unable to overcome the extremely small number of clusters. For Cameron and Miller (2015) and Webb (2014), six clusters seems to be the lower bound for numerical methods to work correctly. Our tests with numerical methods on four and two clusters do not return p-values of reasonable size, given the magnitude of the coefficients and the number of observations in the data.

The reason why these clustering approaches made your p-values explode is because you cannot make statistical claims about country level differences because there are only four countries and outcomes and standard errors are correlated within countries. It would be like running an experiment with one (pre-existing) group assigned to treatment and the others assigned to control. When you cluster standard errors at the level of treatment, which you should, you do not find statistically significant results.

I think that when it comes to making comparisons across countries you should just show the mean values and describe why statistical tests don't work.

The preceding discussion about clustered standard errors should not draw away from the fact that we believe Huber-White robust standard errors are the appropriate adjustment to be made for our data and population of interest. We present the preceding discussion in the interest of completeness, transparency, and to allay any concerns the Reviewer may have that we did not take these concerns seriously.

Some regressions include one other dummy variable plus country fixed effects (Tables S11, S12), survey round and country fixed effects (S13), consumption quintile and country fixed effects (S27, S29), or educational activity and food insecurity (S30, which is a very strange regression to run). Why? Is country correlated with the dummy variable of interest in these cases?

Yes, the Reviewer is correct, we control for country in these regressions because we are concerned about correlation between country and the dummy variable. This applies to the rural/urban dummy (since each country has different degrees of urbanization), to concerns (since disease prevalence varies by country), and to round (the timing of each round differs by country). This also applied to gender of household head, though we have now dropped all of those regressions per the Reviewer's suggestion.

The reviewer is correct that this logic does not apply when we are testing for differences in consumption quintile. Consumption quintiles are calculated within each country and thus are uncorrelated with country fixed effects. We have now dropped country fixed effects from these regressions (S17, S23, and S25). As one would expect, since country fixed effects are uncorrelated with quintiles, the results do not change when we drop country fixed effects.

The bottom line here is that I think the authors could do most of the meaningful comparisons with simple t-tests, considering the explanatory variables of interest are all binary. In some cases, it may be important to do multivariate regressions and/or use interaction terms, in which case econometric analysis makes sense. In these cases, the authors need to provide a better motivation for doing this more complex analysis, because as the paper is now, I do not think these analyses bring much insight, e.g. gender of household head x survey round in S16 and S25, or using school activity as an explanatory variable for food security (S30).

Yes, the Reviewer is correct, in several cases the regression results provide statistically identical results to simple t-tests. However, in many cases, the p-values change when we calculate the Huber-White standard errors.

These standard errors are not appropriate.

We believe that using the multivariate regressions to test differences is superior to simple t-tests for two reasons. First, correcting for heteroskedasticity does matter in our context. Second, we can calculate the same number of tests much more efficiently using multivariate regression, as the regression allows for testing pairs simultaneously, as opposed to the sequential nature required with t-tests.

You can run a regression without any kind of clustering or fixed effects that is essentially identical to using multiple t-tests. This is not a good argument for using multivariate regression.

I also want to raise the issue of using female headship as an explanatory variable. What is the authors' motivation for doing this? Female headship is correlated with many variables that the authors do not control for in any way. So, while it can be a predictive variable, it is not safe at all to attribute any differences in outcomes to the gender of the household head, all else equal. Given this and given that the authors do not propose any policy remedies conditional on gender

of the household head, I do not see a point in doing this analysis. If the authors do want to include it they need to caveat the findings appropriately.

We had included information about female headship based on feedback from presenting results at workshops and seminars. That said, we agree with the reviewer that female headship is likely correlated with a number of unobservables. This means, as the Reviewer points out, that we cannot attribute all of the difference in headship to gender. Because of this, we have removed all tables, figures, and discussion relating to female headship.

Some other issues:

Page 25: "Figures S2 and S2..."

We have removed Figure S2 so this is no longer an issue.

It is unclear why some results were selected to be put in figures in the main text whereas others were relegated to Supplementary Materials. I think the authors should be far more judicious about what tables and figures to put in the paper at all, and about which ones to put in the main text. These should be the ones with the key take-aways. Figure S2 has food security on the horizontal axis, but this is not mentioned at all in the paper when discussing Figure S2 (page 9). I think S2 can be removed.

Per the Reviewer's suggestion, we have removed the supplementary figures in order to clarify the key takeaways and remove the results pertaining to the gender of the head of household.

Author Rebuttal, second revision:

Response to Editor

In what follows, we reproduce (in *italics*) and respond to each of your comments.

You will see from their comments that Reviewers 1 and 3 highlight remaining concerns about the analyses and interpretations of results in your manuscript. Reviewer 3 in particular indicates that their previous concerns about the methodological approach and the handling of standard errors have not been addressed appropriately in this revision. In order to identify the most appropriate resolution for these issues, we submitted your manuscript to our reviewer consultation process, in which we sought Reviewer 1's feedback on Reviewer 3's outstanding concerns. Reviewer 1 noted that, although reporting comparisons of means with t-tests would result in a simpler analytical approach, your existing approach is appropriate; and that your treatment of standard errors is appropriate, as, on the assumption that the samples in the four countries are each drawn with equal probability of selection, there is no non-independence of sample units within countries. You will find Reviewer 1's full feedback included with the reviews at the end of this letter. Based on this feedback, we ask that you retain your regression analyses and treatment of standard errors as they currently stand.

Thank you for the clarification on how to reconcile the comments from Reviewers 1 and 3. We have retained the current regression analyses and treatment of standard errors.

Although we do not require country-clustered standard errors or a change in overall analytical approach, we do require that you thoroughly address all other reviewer points, including the need for population weighting, concerns about the strength of the food insecurity analysis, and the need to avoid causal claims and implications.

We have provided point-by-point responses to all Reviewer comments below.

Finally, your revised manuscript must continue to comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a copy of our checklist that lists all of our requirements. I have also attached a template manuscript file that exemplifies our policies and formatting requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

We have worked to bring the manuscript in line with editorial policies per the checklist provided.

Response to Review #1

In what follows, we reproduce (in *italics*) and respond to each of your comments.

The main statistical issue relates to the estimates of quantities across the 4 countries, which are an important part of the paper (to take one example, the 61% of adult population suffering moderate or severe food insecurity L173). It is not clear whether these estimates have been weighted by country population size, I could see no reference to this in the paper. If the authors have not applied such a weight then these estimates will over-represent the populations in the smaller countries like Malawi and under-represent the populations in the larger countries like Nigeria. The sample sizes of the surveys have not been drawn proportionately to the sizes of the population in each country so this needs to be done via weighting at the analysis stage. Consider a global survey of all countries where the sample size in each country is approximately 1000. If a global average is produced by taking the unweighted arithmetic mean across all countries then an individual in China will be given equal weight to an individual in Iceland, despite the vast difference in the sizes of the respective populations.

We apologize for the confusion: all figures and regressions use post-stratification adjustments that ensure that the population estimates implied by the sampling weights in each country match with the official population projections for that country at the time of the phone surveys. We have clarified this in the Results section. Further, more details are provided on post-stratification adjustments as well in the Methods section.

Second, while the revised manuscript is now more careful about claims of causal impact of covid 19 on the survey variables, there are still problems with how this is treated. Sometimes this is simply the choice of language, for example L105 refers to 'impacts' of covid. In other places it is more fundamental. For example, there is a problem with assuming that all lost income since March is due to the pandemic. At the very least this must be considered an upper bound. That is to say that some households would have lost income from some sources between March and the time of the survey in the absence of the pandemic so we cannot attribute all lost income since March to coronavirus.

The Reviewer's point is well taken. We have removed causal language from the specific cases the Reviewer mentions and have revised the manuscript throughout with an eye to removing any other references to causality.

Elsewhere, breaking down food insecurity estimates by pre-covid consumption quintiles is presented as a mitigation of the problem of not having pre-covid estimates of food insecurity. It is not and should not be presented as such. The authors are correct that levels of food insecurity are of interest in and of themselves but they are not able to produce estimates of change in food insecurity for three of the four countries and presenting the estimates by consumption quintiles does not change this.

We apologize for the confusion, it was not our intention to imply that using pre-COVID consumption quintiles resolved issues of causality. We have revised the section on food insecurity in order to ensure that readers will not misinterpret the use of quintiles as such.

As per the comment above, I am unconvinced that the estimates presented from L230 can be attributed solely to the pandemic. Presumably there would be a non-zero % of people in these countries who cannot access medicines, purchase staples and so on in pre-covid times. Greater care is therefore needed in how these estimates are presented and interpreted.

We have revised this section to address these concerns.

I am not persuaded by the claim in the discussion L208 that the analysis provides insight on the effects of government policies. Where is any link made in the analysis between variation in government policy between countries and the estimates presented? The only place I can see that this is done is where it is argued that awareness of government actions and precautions against infection in Malawi is lower because the government's lockdown policy was overturned in the courts. But this linkage between policy and perception is, at best, weak. Stronger evidence of a link between the estimates presented and variation in government policy is needed or this claim should be removed.

It was not our intention to make the argument that the Reviewer draws attention to. We agree with the Reviewer that such an argument would be unconvincing. The objective of the sentence was to catalog the broad topics (including awareness of government actions) discussed in the paper. We have written the sentence accordingly to clarify our objective.

The authors should be more careful about referring to lockdown restrictions as causing economic problems (e.g. 'government efforts to limit the spread of SARS are having profound economic impacts' l19). It is clear that the immediate consequences of lockdowns are economically damaging but the counterfactual here is not normal life and there are good reasons to think that the economic consequences of taking no action would be considerably worse.

We agree with the Reviewer and have removed all language stating that lockdown restrictions cause economic problems.

Some explanation of why these four countries were selected would be helpful, even if it was just because a sampling frame in the form of pre-covid surveys was only available for these countries.

We have included more details about the sampling frame in the Methods section. Specifically, the selection of these countries was influenced by the public availability of unit-record survey data at the time of the initiation of our work. Since this work began during the summer of 2020, more phone survey rounds have been made publicly available for these four countries, as well as for additional countries including Burkina Faso, Chad, Djibouti, Georgia, India, Kenya, and Mali.

L38 'household individual and child levels'. Households and individuals are different levels of analysis but children also constitute an individual level element.

We have clarified phrase throughout the manuscript these phrases which now refer to "household- and individual-levels, including adult- and child-levels".

L65 "we use well established sampling techniques" should presumably refer to weighting adjustments not sampling techniques (this phrase is repeated in the supplementary appendix).

The Reviewer is correct and we have adjusted this phrase accordingly.

L80 what does 'near real-time' mean?

What we intended with the phrase "near-real-time" is to highlight the fact that the time lag between the data collection and our analysis is much shorter than standard empirical microeconomic studies. We recognize that the phase is colloquial and have reworded the sentence to provide more specificity.

L125 'lack of' should be 'lower'

We have changed this phrase.

Chart D should be presented using single bars indicating % correct rather than separate bars for true and false.

We have revised Figure 1D in line with the Reviewer's suggestion.

As noted above, no reason is suggested for the lower levels of knowledge/awareness in Malawi but the section heading suggests that it is due to 'government action'. This heading would be more accuarte as 'awareness and knowledge of covid-19' or similar.

We agree with the Reviewer and have revised the section title to now be "Awareness and knowledge of COVID-19 are reflected in household behaviour".

Figure 2(d) is unclear. Presumably the %s of yes and no responses sum to 100% so why show bars for yes and no? Only a bar for yes is needed. 'Moderate or severe' and 'moderate' – are these groups independent or can the same respondent be in both groups?

We recognize that Figure 2D was not as clear as it could have been. We have worked to address this in the figure and in the text. The X-axis is not percentages of yes and no, but represents the prevalence of moderate and/or severe food insecurity, from FIES. The FIES measure is a probability that a household is either (1) moderately or severely food insecure or (2) severely food insecure. As such, the bars in the figure, do not report percentages of yes and no, but rather the prevalence of each category of food insecurity. The bars are divided into yes and no categories in order to demonstrate that people answering "yes" indicating that they are concerned" are also more likely to be food insecure.

The reference to student teacher contact is ambiguous – what constitutes contact here and over what time frame?

We have provided more details in the revised manuscript to clarify what is included with the measure of student-teacher contact. The reference period is one week prior and includes activities in a variety of formats, such as SMS, online applications, email, mail, telephone, and/or WhatsApp.

In the discussion I do not consider the fact that there is a lag between data collection and reporting of survey findings to be a particular limitation, nor is it a limitation that these data do not provide a complete account of the pandemic in these countries. These data provide a snapshot of what was happening at this particular point in time, that is not a limitation. It would be better to use this space to discuss the more significant limitations relating to the cross-sectional nature of much of the evidence and the difficulty of linking the estimates presented to the impact of the pandemic.

We agree with the Reviewer that these "limitation" to the completeness of the data are not really limitations to the analysis in the paper. We still think it is important to point out that the data collection process is incomplete and analysis lags behind data collection by several months. In the revised manuscript we now refer to these as "considerations." We have added to these considerations a short paragraph that addresses the limits of our analysis, as the Reviewer points out.

I found Extended data table 1 giving the sample sizes and response rates of the surveys in each country to be unclear. Some explanation is needed for why the number of attempted interviews is less than the number of households available on the frame and why this varies across rounds. Additionally, I assume that the response rates to the pre-covid surveys were less than 100% so the net response rates for each round should account for the initial nonresponse.

We have revised the note in the table to clarify further. Specifically, the rows in the table report the response rate, number of attempted interviews, and number of completed interviews for each country in each round. The total number of households in the pre-COVID-19 surveys is reported in the bottom row, which does not vary by round. This figure also includes households that do not have any phone contact information and thus are outside the scope of the phone survey. The response rate is calculated as number of completed interviews divided by the of attempted interviews. The number of attempted interviews is declining over time since the surveys do not attempt to recontact households that refuse to be interviewed in a given round.

Response to Review #3

In what follows, we reproduce (in **bold**) only the comments arising in the second round of review and our responses to them.

Unfortunately, the paper has not improved much since the first submission. It still uses unnecessarily complex and incorrectly applied econometric methods when t-tests would be sufficient to get most of the points across. I cannot tell if the results in the main figures (those not in the supplementary material) are even using those regression results, or if they are just presenting descriptive statistics (I see no confidence intervals here). If the econometrics are only being used for results presented in the supplementary material, why use them at all?

We regret that we were unable to satisfy the Reviewer on these issues. The Editor has requested that we retain the current analysis and so our responses focus on the other comments.

The results in the main figures are population weighted estimates of the share of people responding to the relevant survey question. Sometimes this is a yes/no response, other times it's a "higher/same/less" response. In order to keep the images as simple and tidy as possible, we did not put error bars on them. Instead, anytime where we discuss differences that are significant or not-significant we provide the statistical evidence in the referenced tables with p-values and 95% confidence intervals. For readers of *Nature Human Behaviour* we believe this is the most straightforward and transparent approach. The reader can quickly see the magnitude of the response for each variable and, if interested, can verify where differences are or are not significant in the tables in the Supplementary Material.

I like this paper as a timely snapshot of how COVID is impacting people in these four African countries. I do not like this paper as an econometric analysis of differences in impacts across countries. The econometrics is useful for some of the analysis described in the paper (specifically, differences across consumption quartiles). But other than that, I really do not see the need to use it and I think the description of the econometric methodology takes up far too much valuable space in what is a very short but dense paper.

Per the Editor's instructions, we have retained the analysis as is. However, to try and accommodate the Reviewer's request as much as possible we have shortened the econometric estimation section, removing the equations for the CEF and OLS estimator.

The organization of this paper is very strange. I realize papers for Nature Human Behavior and other such journals differs from the structure of papers in economics journals. The structure of this paper seems to be a mix of the two. For instance, the "Results" section starts off with two paragraphs about sampling and two paragraphs about data collection. These are not results; they belong with the methods.

We agree that the structure of *Nature Human Behaviour* articles appears strange to us as economists. Working with the Editor, we have done our best to bring the paper in-line with the editorial policies. A component of the structure of articles in the journal is that the main text should only contain three sections (Introduction, Results, & Discussion). Methods come at the end. Because of this structure, it is common for the second section of the paper (Results) tostart with a brief summary of the method to provide context for the reader Gallotti et al. (2020) in NHB provides an example of this. We have followed these guidelines to the best of our ability.

Given how much is already packed into this short paper, I disagree that there is value knowing the status of food insecurity during the pandemic without knowing the status of food insecurity before the pandemic. You can easily make the paragraph starting on line 170 about Nigeria only, because that is the only country for which you have appropriate data to make this analysis. Looking at food insecurity by consumption quartile offers no further insight on the role of COVID: under any circumstances you would expect more food insecurity among those with lower consumption. Heterogeneity in food security across countries is meaningless for this analysis. The evidence on Nigeria is striking, and throwing in this other stuff dilutes this part of the paper.

We appreciate the Reviewer's perspective that the food insecurity information is less interesting in Ethiopia, Malawi, and Uganda when compared to Nigeria. However, Reviewer 1 writes "levels of food insecurity are of interest in and of themselves." In trying to balance these conflicting perspectives, we have decided to retain the results. Our justification is that not every reader will find every result in a paper interesting. But, if some readers do find a result interesting, that result should be retained, and the disinterested reader can skip or ignore it. While we understand that this might not satisfy the Reviewer, we trust the Reviewer can understand and appreciate the difficulty in balancing conflicting recommendations from multiple Reviewers.

I suggest "We then use econometric methods common to the field…" or "We then use common econometric methods…". "Simple" is vague. Along these lines, there is no reason to present the OLS estimator in this paper. Lines 527-544 are a waste of valuable space. Anyone who knows

anything about econometrics knows this estimator. For everyone else, this passage will be completely meaningless.

We appreciate this suggestion and have adopted the Reviewer's phrasing. We have also eliminated the derivation of the OLS estimator.

I stand by this. The paper would be better if the econometrics were mostly if not totally removed and it was written as a descriptive paper. The econometric methodology certainly gets much more attention than it deserves in the paper given it is (a) very simple, and (b) in most cases not necessary beyond a t-test.

Per the Editor's request, we have retained the current econometric methodology but have shortened the discussion of it in the Methods section.

This is the case with the LSMS data, is it not? If it is, then even if you make the argument that clustering should be informed by sample construction and not the level of treatment, then you need to cluster at the village level rather than use Huber-White robust standard errors.

The above comment, and the remaining comment below pertain to which robust standard errors are appropriate and whether the econometric methods used in the paper are appropriate. Per the Editor's request, we have retained the EHW robust standard errors and the econometric method.

This argument is incorrect. You do not have random samples of individuals taken from the populations of these countries. There is multi-level sampling, which means that you would need to cluster at the village level if you are determining your clustering based on sampling. But the reason you need to cluster your standard errors at the country level when you estimate country fixed effects as your outcomes of interest is that treatment is "assigned" at the country level. This point is clear on page 1 of the Abadie et al. (2017) paper: "Instead of a sampling issue, clustering can also be an experimental design issue, when clusters of units, rather than units, are assigned to a treatment". In this paper you make claims about outcomes being different in Malawi because Malawi had different policies about COVID. Thus, the treatment is assigned at the country level. Thus you need to cluster at the country level.

Statistical claims about country level differences because there are only four countries and outcomes and standard errors are correlated within countries. It would be like running an experiment with one (pre-existing) group assigned to treatment and the others assigned to control. When you cluster standard errors at the level of treatment, which you should, you do not find statistically significant results.

I think that when it comes to making comparisons across countries you should just show the mean values and describe why statistical tests don't work.

You can run a regression without any kind of clustering or fixed effects that is essentially identical to using multiple t-tests. This is not a good argument for using multivariate regression.

Decision Letter, third revision:

18th February 2021

Dear Jeff,

Thank you for submitting your revised manuscript "Socioeconomic impacts of COVID-19 in low-income countries" (NATHUMBEHAV-201012983C). It has now been seen by one of the original referees, and their comments are below. As you can see, the reviewer finds that the paper has improved in revision. We will therefore be happy in principle to publish it in Nature Human Behaviour, pending revisions to satisfy the referee's final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Please do not hesitate to contact me if you have any questions.

Sincerely,
Aisha

Aisha Bradshaw
Editor
Nature Human Behaviour

Reviewer #1 (Remarks to the Author):

One small point. In the conclusion the authors acknowledge a limitation of their study as its primarily cross-sectional nature. Weighting is then noted as a mitigation strategy for making causal inferences with observational data. This is not really accurate - the weighting used in the analysis is primarily intended to reduce bias due to nonresponse (of various kinds). there is indeed a real risk that some of the estimates in the paper are subject to nonresponse bias and it is appropriate to acknowledge this and that weighting may not completely correct for this.

natureresearch

Decision letter, final requests:

22nd February 2021

Dear Dr. Michler,

Thank you for your patience as we've prepared the guidelines for final submission of your Nature Human Behaviour manuscript, "Socioeconomic impacts of COVID-19 in low-income countries" (NATHUMBEHAV-201012983C). Please carefully follow the step-by-step instructions provided in the personalised checklist attached, to ensure that your revised manuscript can be swiftly handed over to our production team.

We hope to receive your revised paper, with all of the requested files and forms, within 10 days. If you anticipate delays, we would be grateful if you could contact us to provide us with an estimate regarding when you will submit these files.

When you upload your final materials, please include a point-by-point response to any remaining reviewer comments.

If you have not done so already, please alert us to any related manuscripts from your group that are under consideration or in press at other journals, or are being written up for submission to other journals (see: https://www.nature.com/nature-research/editorial-policies/plagiarism#policy-on-duplicate-publication for details).

Nature Human Behaviour offers a Transparent Peer Review option for new original research manuscripts submitted after December 1st, 2019. As part of this initiative, we encourage our authors to support increased transparency into the peer review process by agreeing to have the reviewer comments, author rebuttal letters, and editorial decision letters published as a Supplementary item. When you submit your final files please clearly state in your cover letter whether or not you would like to participate in this initiative. Please note that failure to state your preference will result in delays in accepting your manuscript for publication.

In recognition of the time and expertise our reviewers provide to Nature Human Behaviour's editorial process, we would like to formally acknowledge their contribution to the external peer review of your manuscript entitled "Socioeconomic impacts of COVID-19 in low-income countries". For those reviewers who give their assent, we will be publishing their names alongside the published article.

Cover suggestions

As you prepare your final files we encourage you to consider whether you have any images or illustrations that may be appropriate for use on the cover of Nature Human Behaviour.

Covers should be both aesthetically appealing and scientifically relevant, and should be supplied at the best quality available. Due to the prominence of these images, we do not generally select images featuring faces, children, text, graphs, schematic drawings, or collages on our covers.

We accept TIFF, JPEG, PNG or PSD file formats (a layered PSD file would be ideal), and the image should be at least 300ppi resolution (preferably 600-1200 ppi), in CMYK colour mode.

If your image is selected, we may also use it on the journal website as a banner image, and may need to make artistic alterations to fit our journal style.

Please submit your suggestions, clearly labeled, along with your final files. We'll be in touch if more information is needed.

Nature Human Behaviour has now transitioned to a unified Rights Collection system which will allow our Author Services team to quickly and easily collect the rights and permissions required to publish your work. Approximately 10 days after your paper is formally accepted, you will receive an email in providing you with a link to complete the grant of rights. If your paper is eligible for Open Access, our Author Services team will also be in touch regarding any additional information that may be required to arrange payment for your article.

Please note that you will not receive your proofs until the publishing agreement has been received through our system.

For information regarding our different publishing models please see our Transformative Journals page. If you have any questions about costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com.

Please use the following link for uploading these materials:
[REDACTED]

If you have any further questions, please feel free to contact me.

Best regards,

Aisha Bradshaw
Editor
Nature Human Behaviour

Reviewer #1:
Remarks to the Author:
One small point. In the conclusion the authors acknowledge a limitation of their study as its primarily cross-sectional nature. Weighting is then noted as a mitigation strategy for making causal inferences with observational data. This is not really accurate - the weighting used in the analysis is primarily intended to reduce bias due to nonresponse (of various kinds). there is indeed a real risk that some of the estimates in the paper are subject to nonresponse bias and it is appropriate to acknowledge this and that weighting may not completely correct for this.

Author Rebuttal, third revision:

Response to Reviewer #1

In what follows, we reproduce (in *italics*) and respond to each of your comments.

One small point. In the conclusion the authors acknowledge a limitation of their study as its primarily cross-sectional nature. Weighting is then noted as a mitigation strategy for making causal inferences with observational data. This is not really accurate - the weighting used in the analysis is primarily intended to reduce bias due to nonresponse (of various kinds). there is indeed a real risk that some of the estimates in the paper are subject to nonresponse bias and it is appropriate to acknowledge this and that weighting may not completely correct for this.

We take the reviewer's point. We have revised the section to read:

These corrections do not in themselves establish a causal relationship between the onset of the pandemic and our outcomes of interest, and there may be remaining selection and non-response bias that the survey sampling weights and econometric methods fail to fully account for. Many adults in these countries were food insecure prior to the pandemic, and the loss of income or revenue may be due to non-pandemic related events. More work is needed to align the post-COVID phone survey data with the pre-COVID face-to-face data in order to establish a pre- and post-pandemic panel.

Final Decision Letter:

Dear Jeff,

I am happy to inform you that your Article "Socioeconomic impacts of COVID-19 in low-income countries", has now been accepted for publication in Nature Human Behaviour.

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see http://www.nature.com/nathumbehav/info/gta). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

natureresearch

For manuscripts submitted prior to 1 January 2021, Nature Research allows authors to self-archive the accepted manuscript (the version post-peer review, but prior to copy-editing and typesetting) on their own personal website and/or in an institutional or funder repository where it can be made publicly accessible 6 months after first publication, in accordance with our self-archiving policy. Please review our self-archviing policy for more information.

Several funders require deposition the accepted manuscript (AM) to PubMed Central or Europe PubMed Central. To enable compliance with these requirements, Nature Research therefore offers a free manuscript deposition service for original research papers supported by a number of PMC/EPMC participating funders. If you do not choose to publish immediate open access, we can deposit the accepted manuscript in PMC/Europe PMC on your behalf, if you authorise us to do so.

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Human Behaviour as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Human Behaviour logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

You will not receive your proofs until the publishing agreement has been received through our system.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

We look forward to publishing your paper.

With best regards,
Aisha

Aisha Bradshaw
Editor
Nature Human Behaviour

P.S. Click on the following link if you would like to recommend Nature Human Behaviour to your librarian http://www.nature.com/subscriptions/recommend.html#forms

** Visit the Springer Nature Editorial and Publishing website at www.springernature.com/editorial-and-publishing-jobs for more information about our career opportunities. If you have any questions please click here.**