## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Protocol for a longitudinal study examining the trajectory of COVID-19, Post-COVID, multidimensional disadvantage and Health Related Quality of Life in India – the IndiQol Project |
| **AUTHORS** | Sinha, Kompal; Gutacker, Nils; Gu, Yuan-Yuan; Haagsma, Juanita; Kumar, Kaushalendra; Aghdaee, Mona |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Cohen, Rubin<br>Syracuse VA Medical Center, Medicine |
| **REVIEW RETURNED** | 12-Dec-2023 |

| | |
|---|---|
| **GENERAL COMMENTS** | The planned study is very interesting and necessary. I do agree that the data generated may serve as a base for future studies. There are a few points:<br>-The authors should comment on limitations of the study, for example much depends on recall of events that took place 3-4 years earlier. On page 21, under protocol, the authors do mention two points:<br>• The survey will collect self-reported data which can result in measurement error due to self reporting bias.<br>• The survey will be conducted across four waves which would result in some attrition in later waves of the survey.<br>These points, and others, may be better if included under a paragraph labeled limitations<br>-Given the economic disadvantage of some household, will they all have telephones? I gather if the household has no phone, this household will not be included, this may bias results somewhat.<br>-From Table 1, indicators for the State of Maharashtra appear better than others, does this impact on survey ? |

| | |
|---|---|
| **REVIEWER** | Avsar, Tuba Saygin<br>NICE |
| **REVIEW RETURNED** | 20-Dec-2023 |

| | |
|---|---|
| **GENERAL COMMENTS** | I'd like to thank the authors for this important work. I've enjoyed reading the protocol, which was very detailed and clearly written. I've made some small suggestions to improve the reporting.<br>The only key issue is to clarify the study aim: is it "to evaluate the socioeconomic impact of COVID-19 and long COVID in India" as stated in the abstract or "to evaluate the impact of COVID-19 and long COVID on HRQoL and its relationship with socioeconomic factors in India"? For instance, in regarding income, the former would cover loss in income due to COVID while the latter is about how the HRQoL impact of COVID differs by income.<br>Abstract |

| | |
|---|---|
| | Line 31: The objective should be updated based on the authors' consideration of my comment above.<br>Introduction<br>Page 3, Line 33: "Of" the people who survive…<br>Aim of the research<br>Page 4, Line 3: I think there is a missing "and" here: "between of long COVID health related quality of life".<br>Methods<br>Page 4, Line 22: There is a typo here: "under-5 years' age".<br>Page 4, Line 29: Citation style seems wrong.<br>Page 4, Line 34-35: Is the population growth relevant to the study outcomes? It is not included in Table 1, so I wonder if it is needed in the text.<br>Page 4, Line 38-43: Is there a simpler way to communicate this information? What does this ratio tell us about the difference between the states?<br>Page 4, Line 39: I think this is the first time DALYs is used; so, should be written in full.<br>Page 4, Lines 45-46: These sentences are not needed as the information is repeated at the end of the next section. I think it fits there better.<br>Page 5, Line 21: Citation needed here for "the norm for survey data collection in household sample surveys in India".<br><br>Page 5, Line 38: Suddenly the IndiQol project is mentioned. It is a bit confusing for the reader as to whether it refers to another project or the current one. Maybe the project name should come earlier or be omitted.<br><br>Table 1: There are superscripts; so, are explanations beneath the table missing?<br><br>Survey timeframe<br>This section mostly is about how the survey will be conducted and by whom. So, I think the title should be changed to reflect that.<br><br>Page 7, Line 5: Should years need to be updated? Has the data collection started? |

## VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Rubin Cohen, Texas Tech University Health Sciences Center Foster School of Medicine

Comments to the Author:
The planned study is very interesting and necessary. I do agree that the data generated may serve as a base for future studies.
There are a few points:
-The authors should comment on limitations of the study, for example much depends on recall of events that took place 3-4 years earlier. On page 21, under protocol, the authors do mention two points:
• The survey will collect self-reported data which can result in measurement error due to self reporting bias.
• The survey will be conducted across four waves which would result in some attrition in later waves of the survey.

These points, and others, may be better if included under a paragraph labeled limitations

Response: Thanks for your comments. We have included a separate section on limitation in the revised manuscript (see page 17). Please note, we have replaced long-COVID with post-COVID throughout the text to follow the current norm.

Comment: Given the economic disadvantage of some household, will they all have telephones? I gather if the household has no phone, this household will not be included, this may bias results somewhat.

Response: Thanks for this comment. The survey will be conducted face to face and does not require for the household to have a telephone to conduct the survey. Although, the respondents are asked to share their telephone number, the purposes of scheduling interviews in subsequent waves. We have clarified this in on page 8 of the revised manuscript.

Comment: From Table 1, indicators for the State of Maharashtra appear better than others, does this impact on survey?

Response: The survey has selected states to represent the distribution of wellbeing indicators across the country. Maharashtra has higher indicators as it is the largest economy in India accounting for 15% of India's GDP, and has been included to represent the economically advanced states. We do not anticipate the survey would not be impacted.

Reviewer: 2
Dr. Tuba Saygin Avsar, NICE

Comments to the Author: I'd like to thank the authors for this important work. I've enjoyed reading the protocol, which was very detailed and clearly written. I've made some small suggestions to improve the reporting.
The only key issue is to clarify the study aim: is it "to evaluate the socioeconomic impact of COVID-19 and long COVID in India" as stated in the abstract or "to evaluate the impact of COVID-19 and long COVID on HRQoL and its relationship with socioeconomic factors in India"? For instance, in regarding income, the former would cover loss in income due to COVID while the latter is about how the HRQoL impact of COVID differs by income.
Abstract
Line 31: The objective should be updated based on the authors' consideration of my comment above.

Response: Thanks for your comment. The objective of the study is to evaluate the how COVID-19 and long/post COVID has impacted HRQoL and the mediating role of socioeconomic factors in India. The survey will capture how COVID-19 has affected the social, economics, and health related quality of life in India. We have revised the abstract for better clarity considering your comment.

Introduction
Page 3, Line 33: “Of” the people who survive…

Aim of the research
Page 4, Line 3: I think there is a missing “and” here: “between of long COVID health related quality of life”.
Methods
Page 4, Line 22: There is a typo here: “under-5 years’ age”.
Page 4, Line 29: Citation style seems wrong.
Page 4, Line 34-35: Is the population growth relevant to the study outcomes? It is not included in Table 1, so I wonder if it is needed in the text.
Page 4, Line 38-43: Is there a simpler way to communicate this information? What does this ratio tell us about the difference between the states?
Page 4, Line 39: I think this is the first time DALYs is used; so, should be written in full.
Page 4, Lines 45-46: These sentences are not needed as the information is repeated at the end of the next section. I think it fits there better.
Page 5, Line 21: Citation needed here for “the norm for survey data collection in household sample surveys in India”.
Page 5, Line 38: Suddenly the IndiQol project is mentioned. It is a bit confusing for the reader as to whether it refers to another project or the current one. Maybe the project name should come earlier or be omitted.

Response: Thanks for your comments and suggestions. We have incorporated all editorial suggestions in the revised manuscript. The IndiQol study appears in the title of the paper, we have addressed your comment and now introduced the project name earlier on in the manuscript.

Table 1: There are superscripts; so, are explanations beneath the table missing?

Response: The superscripts in Table 1 are now explained in the table footnotes.

Survey timeframe
This section mostly is about how the survey will be conducted and by whom. So, I think the title should be changed to reflect that.

Response: We have changed the title for the subsection “survey timeframe” to “survey design”.

Comment: Page 7, Line 5: Should years need to be updated? Has the data collection started?

Response: Thanks for this comment. The survey timeframe is correct in the paper. The survey has been piloted and data collection is underway.

Reviewer 3:
Paula Lorgelly, University of Auckland

This protocol describes the establishment of the IndiQol study, an examination of long COVID, multidimensional disadvantage and HRQoL in India. I understand the project is already in the field

collecting data, so none of my comments fundamentally change the proposed plan, but I believe they can improve the paper as it is currently written.

Comment: The need for the study is well introduced, India as a country experienced considerable burden with respect to the COVID-19 pandemic, and therefore it is interesting to understand the impact of deprivation and on HRQoL on long term impacts. However, some of the statistics presented are confusing. The opening sentences of the second paragraph suggests that 59% of those surviving COVID have long COVID (persistent symptoms). In fact it is 59% of individuals with long COVID (the sample of the cited study) that have poor quality of life. This and the other statistics, including 37% of individuals with COVID-19 having long COVID, overstate the issue. It is now accepted, given vaccination and antivirals, it is thought that 6-10% of those with COVID-19 go on to develop long COVID. This estimate is confirmed for India https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9874641/

Response: Thanks for your comment. We have revised the statistics in the paper now.

Specifically, the following text has been added to the text:
"Of patients infected with COVID, post-COVID is estimated to occur in 10–20% of cases and affects people of all ages, including children, with most cases occurring in patients with mild acute illness.[15,16] While the conservative estimate of post-COVID is 10%, the incidence is estimated to be 10-30% for non-hospitalised cases; 50-70% for hospitalised cases and 10-12% of vaccinated cases.[17] In the case of India, it is estimated that 6-10% of those with COVID go on to develop post COVID.[16]"

Additionally, please note, we have replaced long-COVID with post-COVID throughout the text to follow the current norm.

Comment: The introduction covers both long COVID and multidimensional deprivation, but there is no introduction to the interest in HRQoL. This includes limited information in the introduction about the use of EQ-5D-5L as a measure of HRQoL. In particular there has recently been developed a long COVID specific HRQoL instrument, PC-COS (post covid core outcome set). Note that EQ-5D-5L is not an acronym https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7334333/.
It is not EuroQol five-dimension five level, it is just EQ-5D-5L.

Response: Thanks for your comments. We have incorporated your suggestion in the revised manuscript. The introduction mentions the interest HRQoL and the use of EQ-5D-5L as its measure. We appreciate your suggestion of PC-COS and will consider it for inclusion in the IndiQol project.

Comment: Top of page 4, is there a word missing between long COVID and health-related quality of life, like 'and'? And should the last part of that sentence be *long* COVID? The description of the sample size calculation is very confusing. What is the effect size? Is it the difference in EQ-5D-5L scores? If so 0.1 is not small it is larger than most estimates of the minimal clinically important difference (0.055-0.08). As the effect sizes in other papers are at the individual level does it make a difference that this study is focusing on household recruitment? The sampling strategy text is very confusing. If the purpose of a protocol paper is to replicate the project, this is not possible given how this text is written. Would a diagram of substrata and sub-substrata (I'm assuming what happens when substrata are crossed) be helpful?

Response: Thank you for your comment. In the revised manuscript, we have corrected the typo on page 4.

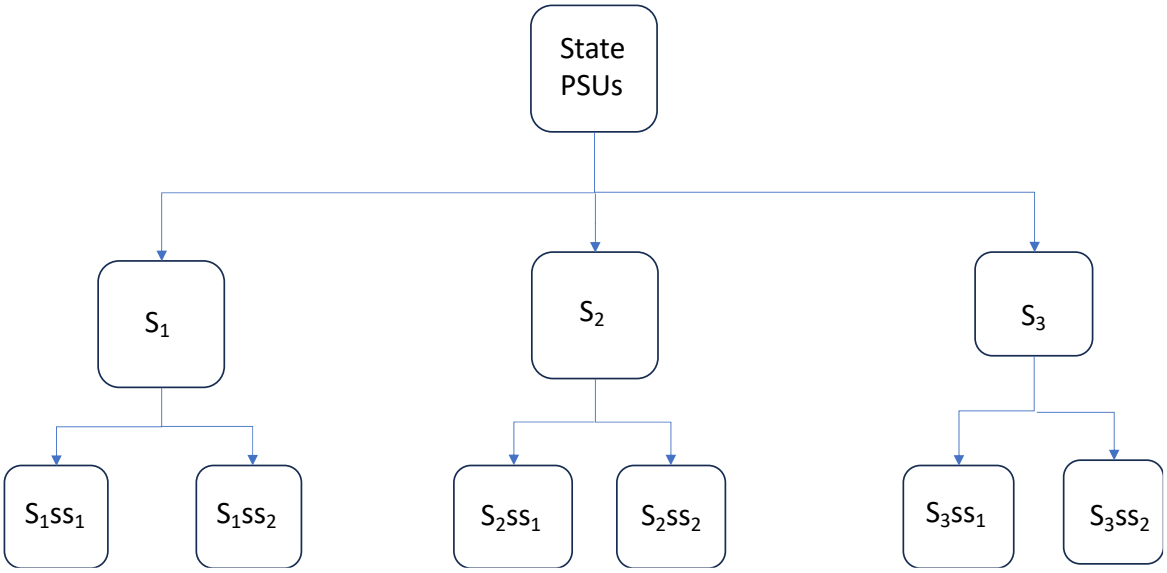
We followed the Cohen's guidelines in making these calculations (Cohen 1988). Effect size is used to evaluate the effect of the study and indicate its significance along with the statistical significance. While P-value talks about the statistical evaluation of research providing information on the presence/absence of an effect, the effect size evaluates the size of the effect. It details on the practical significance of the research outcome with a large effect size meaning that the study has practical significance and small effect size means that the research has low practical significance. Measured as Cohen's d, this study uses effect size to find the sample size required for sufficient power of our study. In deciding the effect size of 0.1 we followed the Cohen's d for standardised effect size of change in quality of life reported by Al Daheri et al (2021) for MENA countries; change in EQ-5D-5L reported by Hegde et al (2021) for India; Choi et. al. (2021) for Hong Kong among others (see section 3.2). For instance, Hegde et al (2021) report dimension wise change in EQ-5D-5L for India in Table 4 on page 501 of their paper. The

mean change (MC) and standardised effect size (SES) in EQ-5D-5L is reported as follows: Mobility -0.08 (MC) and 0.1420 (SES); Self-care 0.06 (MC) and 0.1312 (SES); Usual activity 0.09 (MC) and 0.1888 (SES); Pain/discomfort 0.11(MC) and 0.1872 (SES); Guided by these effect sizes, and the norms used in other population-based surveys such as WHO SAGE study, we have calculated the sample size for our study.

The objective of the sampling design is first to capture the PSUs based on the population size and then to capture the socioeconomic variation. We adopt a multistage stratified sampling approach was adopted for our survey. A stratified sampling design allows us to partition the population into strata based on common sampling characteristics of the population and then selecting samples independently from each stratum. A diagram of the project objective and sampling design is presented below.

A multistage stratified sampling approach was adopted to get our sample of 3000 households. The first stage selects the six states. The second stage considers each state as a sampling frame and stratifies districts into sectors (rural/urban). For each sector the districts are stratified based on probability proportion to size design by using total households in each PSU to create three sub-strata (S1, S2 and S3). Within each of these sub-strata we stratify the PSUs based on proportion of SC/ST population using the median as the cut-off point. Two sub-substrata ($S_1ss_1$ and $S_1ss_2$) comprising of the proportion of SC/ST less than 50% in $S_1ss_1$ and proportion of SC/ST>50% in $S_1ss_2$ are formed. This method of stratifying sub-substrata within a sub-strata is what is referred to as substrata crossing in the text. This sampling approach is followed by the Demographic and Health Surveys i.e., National Family Health Survey (NFHS) for India. Using simple random sampling 30-35 PSUs from each state are selected. The final step of selecting households from the selected PSUs was done using listing of households from a random segment of the PSUs.

| | IndiQol Study |
|---|---|
| Broad objective | Post COVID health-related quality of life (HRQoL) as measured by EQ-5D-5L and mediating role of multidimensional deprivation. |
| Sample domain | Six states of India which represent socio-demographic, health status and infrastructure, economic development variation across states of India |
| PSU/CEB (Cluster) | Random selection of 30 sites (termed "cluster") from within each geographical area for which individual results are desired.<br><br>Stratified selection of PSU/CEB from each of the six selected states. |
| Population / unit of interest | Identification of age group of interest. Random selection of households from the selected PSU/CEB |




IndiQol Sampling Design

Comment: For the survey timeframe, please reflect on the seasons of the 4 monthly intervals, this is likely to be impactful. Is a gift voucher to the value of 2 euros (what is this in rupees?) the norm for field-based interviews in India?

Response: Thanks for this comment. In the revised manuscript we have reflected on the season of the 4 monthly intervals. There is no specific norm of giving a gift voucher for field-based surveys in India. The gift voucher of 2 Euros given at the end of survey is equivalent to around Rupees 175. This amount was decided in discussion with other researchers and field survey experts in India.

Comment: The survey questionnaire section (3.4) is the first mention of pre-pandemic recall, this should be introduced earlier. Are there high rates of long COVID in India? Is there a reference to this?

Response: Thanks for your comment. In the revised manuscript we have introduced the pre-pandemic recall earlier in the text. Long/post-COVID is prevalent in India for almost three years and has been covered in mainstream news outlets (Times of India, The Hindu, The Economic Times). These is social stigma around the disease and not many people are talking about it (The Guardian).

Comment: The statement that most COVID symptoms last at least 3 months is incorrect (page 7, line 38). Most COVID symptoms last up to 2 weeks. By definition 3 months of symptoms is long COVID.

Response: Thank you for this comment. Indeed, after reading the manuscript again, it became clear to us that there is a typo in the statement (page 7, line 38). It should be long-COVID symptoms instead of COVID symptoms. Please note, we have replaced long-COVID with post-COVID throughout the text to follow the current norm. We have addressed this error in the revised manuscript.

Comment: It appears from Table 2 that everyone surveyed will be asked about their COVID experience, so vaccination and infection, but also long COVID symptoms and stigma questionnaires. Will there be some filter? How will individuals identify that they have long COVID? Will the survey interview process provide a diagnosis and if so, what will individuals do with that information?

Response: All individuals will be asked about their COVID experience including vaccination history, infection and COVID related stigma questions. The post-COVID symptoms will be asked only for respondents reporting a COVID history i.e., currently infected with COVID or infected in the past. We have added a table footnote to clarify this confusion. Individuals will not be able to identify whether they have post-COVID, they will be asked if they have had any of the symptoms associated with post-COVID. The survey interview process will not provide any diagnosis to the individuals.

Comment: How are the authors defining the period “during COVID” for financial hardship during COVID? Is this the peak pandemic period so March 2020 to late 2021/early 2022.

Response: Thanks for your comment. The period “during COVID” is indeed referring to the period since the onset of the pandemic i.e., since March 2020.

Comment: I am left wondering if the objective is to look at the long term consequences of the pandemic, of which long COVID is a specific consequence. Unless the proportion of long COVID is a lot higher than an optimistic value of 10% then much of the rich data will be in those without long COVID. Yes some of the sample will have long COVID, but the questions on multiple disadvantage and the impact on HRQoL post-pandemic are equally of interest. I would encourage the authors to expand their focus, issues of long COVID can still be addressed but need not be the defining clinical feature of the IndiQol study.

Response: Thank you very much for this comment and suggestion. Indeed, after reading the manuscript again, it became clear to us that discussion about the objectives was somewhat unclear. We are delighted to read your suggestion as this is indeed the objective of the study i.e., understand the consequences of pandemic along with post-COVID as a specific consequence. This aspect is included in the design of study where we consider a population representative sample rather than a sample of COVID patients only. You raise a valid point that the proportion of long/post-COVID could be low and much of the rich data might be for those without post-COVID. We have carefully considered this issue and have resolved to interviewing all adult members of households. The question on multiple disadvantages and the impact on HRQoL post pandemic will be addressed in this study.

Comment: Please review the use of acronyms and give them in full, e.g. DALYs, CAPI.

Response: We have reviewed the use of acronym and give then in full. Following changes have been made.

References

Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb). 2021 Feb 15;31(1):010502. doi: 10.11613/BM.2021.010502. Epub 2020 Dec 15. PMID: 33380887; PMCID: PMC7745163.

Arjun MC, Singh AK, Roy P, Ravichandran M, Mandal S, Pal D, Das K, Gajjala A, Venkateshan M, Mishra B, Patro BK, Mohapatra PR, Subba SH. Long COVID following Omicron wave in Eastern India-A retrospective cohort study. J Med Virol. 2023 Jan;95(1):e28214. doi: 10.1002/jmv.28214. Epub 2022 Oct 25. PMID: 36224705; PMCID: PMC9874641.

Ledford, H. (2024). Long COVID is a double curse in low-income nations—here's why. Nature, 625(7993), 20-22.

**VERSION 2 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Cohen, Rubin<br>Syracuse VA Medical Center, Medicine |
| **REVIEW RETURNED** | 04-Apr-2024 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors have responded to my concerns as well as the concerns of others quire well. The revised version is greatly improved. I have no issues. I commend the authors on their work to improve this manuscript. |