PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Understanding public preferences and trade-offs for government responses during a pandemic: A protocol for a discrete choice experiment in the United Kingdom
AUTHORS	Genie, Mesfin G; Loría-Rebolledo, Luis Enrique; Paranjothy, Shantini; Powell, Daniel; Ryan, Mandy; Sakowsky, Ruben Andreas; Watson, Verity

VERSION 1 – REVIEW

REVIEWER	Alex Cook National University of Singapore
REVIEW RETURNED	18-Aug-2020

GENERAL COMMENTS	Introduction 6.9 Very Eurocentric references here. The Italian and Spanish measures don't seem particularly stringent compared to how places like Viet Nam, China or the Republic of Korea have controlled their outbreaks. 6.14 Ditto, haven't local lockdowns been implemented outside of "Western" countries? Methods 8.30 Of all the attributes considered, postponing medical care seemed off. Lockdown type and length are public health measures implemented to reduce infection, but postponing therapy in hospital would be to ensure capacity of the healthcare system. Lockdown is political but hospital management is operational. If there was another attribute that was dropped to make way for this, perhaps it's worth reconsidering that decision. 9.51 Should collect sex/gender as well. Consider ethnicity since that's been highlighted as an important risk factor in the UK. Worth asking about whether they know anyone who was infected/hospitalised/died and not just personal experience. 11.22: Where -> where. "And" not italicised. Why is there a small i in the range for j? 12.4 For both sections, confirm with the named people that they do not object to being identified in your paper, in particular your public representative. Tables and Figures T1: Check the colour diagram can be interpreted by those with different degrees of colour blindness? I don't know how to interpret attribute 6. Let's say I have a trolley with goods costing £100 in today's money. In a year's time, the cost of the food will have changed relative to my salary. Perhaps

	my salary went down, or perhaps the food now costs £150. *But I am still able to buy it.* So the answer logically should be 100%, unless I'm down to my last shilling. I don't think this works as a measure of loss of purchasing power. F1 White text on light yellow background does not work well. Consider tabbing the levels, so the attributes in bold and the levels in normal face are both left flush. I think that may make for easier reading. OSM1 I didn't "get" the linkage between the twitter analysis and the development of the attributes and levels. In the word clouds I see meaningless fragments of ideas like 'people' and 'government'. What do these contribute to the work? The bar charts are quite amateur looking. R syntax in the axis, stretched and squashed images of different sizes, unlabelled tick marks… I can't interpret the x axis at all. The entire OSM1 seems like a dump for unfinished ideas.

REVIEWER	Mickael Hiligsmann Maastricht University, the Netherlands
REVIEW RETURNED	21-Sep-2020

GENERAL COMMENTS	In this study, the authors provided the protocol of a discrete choice experiment to understanding public preferences and trade-offs for government responses during a pandemic. The research is important, timely and has been adequately described in this protocol. The type and number of attributes seem also adequate. The research team has further sufficient experience to adequately perform the study. Some suggestions here below to strengthen this article: - In the development of attributes and levels, it is unclear how and when the literature search was performed. More details on the search conducted (including platform used), time period and inclusion criteria would be needed, especially as studies on this topic are quickly becoming available. - Could the authors provide more information on their experimental design and which priors/direction they will use for which parameters? - The authors mentioned that they "will ask respondents how likely they are to comply with the chosen scenario". Could the authors explain how these data will be analysed? - For statistical analysis, it would be nice to know whether some variables will be included as categorical or continuous variable (or maybe both will be tested by the authors?). - It is a little bit unclear how the socio-economic characteristics will be analysed/tested in the model. Further explanation could be interesting. - A latent class model could also provide relevant information and potentially be (at least) tested.

VERSION 1 – AUTHOR RESPONSE

Reply to Reviewer #1

Comment #1.1

6.9 Very Eurocentric references here. The Italian and Spanish measures don't seem particularly stringent compared to how places like Viet Nam, China or the Republic of Korea have controlled their outbreaks.

Reply #1.1

While we acknowledge your point about the Eurocentric references used in the protocol, we focused on European countries mainly for two reasons:

1. We initially developed the attributes and levels based on government guidance and interventions observed in European countries where the infection waves were concentrated (see Table 2). These interventions were the ones that the UK government and press referred to when developing or announcing UK restrictions.
2. We did not refer to the situation in Vietnam or China because the context in these countries (in terms of culture, politics, compliance with rules) is different from European countries. The political responses in these countries were therefore less relevant for the UK context than those of European neighbours and, we feel, less likely to compare to the attitudes of UK-based respondents. Moreover, we recognise these countries had experience in responding to outbreaks of other infectious diseases, including SARS, MERS, measles and dengue.

We have now edited the text in the **Introduction**, on **Page 4,** and the first paragraph reads as follows:

The public health response of governments to the COVID-19 pandemic has differed across countries. Responses have mostly involved lockdown measure that encourages social distancing[1] to slow the spread of the disease.[1,2] The timing and strictness of these measures has differed across countries. Italy and Spain introduced early strict lockdown measures,[3,4] while restrictions in Sweden and the Netherlands were less severe.[5,6] Responses in China and Vietnam were more stringent [7,8] due to prior experiences of responding to outbreaks of other infectious diseases, including SARS, MERS, measles and dengue.[9,10] Policies in the UK shifted from a more relaxed initial approach towards stricter measures as the pandemic progressed.[11] As lockdowns have eased (as of 27th of July 2020), we have seen local lockdowns introduced to deal with spikes in infection rates (e.g. Leicester (England), Melbourne (Australia) and Barcelona (Spain).[12,13,14]

Comment #1.2
6.14 Ditto, haven't local lockdowns been implemented outside of "Western" countries?

Reply #1.2

At the time of writing of the protocol, we were aware of local lockdowns only in Leicester (England), Melbourne (Australia) and Barcelona (Spain). We have now put a timestamp for cases of local lockdowns observed that correspond the time writing this protocol (see above).

[1] Social distancing, also called "physical distancing," means keeping a safe space between yourself and other people who are not from your household.[24]

Comment #1.3

8.30 Of all the attributes considered, postponing medical care seemed off. Lockdown type and length are public health measures implemented to reduce infection, but postponing therapy in hospital would be to ensure capacity of the healthcare system. Lockdown is political but hospital management is operational. If there was another attribute that was dropped to make way for this, perhaps it's worth reconsidering that decision.

Reply #1.3

This is a very good point. Whilst acknowledging the reviewer's comment, we argue that '*postponement of usual non-urgent medical care*' is a relevant attribute in the UK context and is under the purview of the government. For example, the Scottish government cancelled routine procedures in 2020 (https://www.nhsperforms.scot/hospital-data/indicator-hospital?hospitalid=3&indicatorid=28). Further, nearly a million appointments for mammograms have been missed in the UK due to the COVID-19 pandemic (https://breastcancernow.org/about-us/media/press-releases/almost-one-million-women-in-uk-miss-vital-breast-screening-due-covid-19). Further, the UK government provided the funding and resources to build additional hospital bed capacity (https://www.england.nhs.uk/2020/04/nhs-to-build-more-nightingale-hospitals-as-london-set-for-opening/, https://blogs.lse.ac.uk/covid19/2020/08/12/could-nightingale-hospitals-be-the-solution-to-nhs-bed-shortages/, https://www.medscape.com/viewarticle/929778). This makes this a relevant attribute to include in the DCE. Further, we have since extensively tested this using think-aloud and found respondents interpret this correctly.

We have now edited the text for attribute 3 in the **Development of attributes and levels** section, on **Page 6,** to make the rationale for using this attribute, and the description of the attribute "postponement of usual non-urgent medical care" reads as follows:

3. Postponement of usual non-urgent medical care: Governments around the world have cancelled usual medical care to deal with staff shortages and help the healthcare system respond to an expected increase in patients. This attribute is particularly relevant in the UK context and is under the scope of the government. For example, the Scottish government cancelled routine procedures in 2020.[31] Further, nearly a million appointments for mammograms have been missed in the UK due to the COVID-19 pandemic.[32] This, along with the extra capacity built (e.g. Nightingale around England and Louisa Jordan in Glasgow), make this attribute particularly relevant for providing policy advice. This attribute has three levels: all non-urgent, non-pandemic-related procedures postponed; some non-urgent procedures postponed, and all procedures continue as scheduled.

Comment #1.4

9.51 Should collect sex/gender as well. Consider ethnicity since that's been highlighted as an important risk factor in the UK. Worth asking about whether they know anyone who was infected/hospitalised/died and not just personal experience.

Reply #1.4.1

The list of sociodemographic characteristics in the protocol did not list all characteristics that we planned to collect data on. We will collect information on sex and ethnicity, along with many others that we think could explain heterogeneity. We have included this information on **Page 8:**

The survey instrument will also collect information on respondents' socio-demographic characteristics (age, sex, education, ethnicity, economic insecurity, health status, country of residence, experience of COVID-19, etc.).

Reply #1.4.2

We have included the following related questions in the survey to capture both personal and non-personal experience:

a. Have the COVID-19 lockdown measures affected the healthcare you or your loved ones normally receive? (Yes or No)
b. How concerned are you that **you** or **a loved one** could die from COVID-19? (a 5-point Likert scale response from 'not at all concerned' to 'extremely concerned')
c. How concerned are you that **you** or a **loved one** could die because you **could not access healthcare during the COVID-19 pandemic?** (a 5-point Likert scale response from 'not at all concerned' to 'extremely concerned')

Comment #1.5
11.22: Where -> where. "And" not italicised. Why is there a small i in the range for j?

Reply #1.5

Thanks for this; we have changed 'Where' into 'where' and checked the notations and a few adjustments are made on **Page 9** line 22 and 23. We have changed the small '*i*' into 'I' in the range for '*j*'.

Comment #1.6
12.4 For both sections, confirm with the named people that they do not object to being identified in your paper, in particular your public representative.

Reply #1.6

Thanks for this reminder; we confirm that the named people do not object to being identified in the paper.

Comment #1.7
T1: Check the colour diagram can be interpreted by those with different degrees of colour blindness?

Reply #1.7

Thanks for raising this point. None of the diagrams include colour-on-colour text and all of the information transmitted via colour schemes (e.g. colour coding for restriction severity) is also supplied via text (e.g. “green restrictions”, “yellow restrictions” etc.).

Further, we have conducted virtual think-aloud (TA) interviews with colleagues from the University of Aberdeen (n=10), members of our Stakeholder Advisory Group (SAG, n=4) and members of the general public (n=23) recruited via two Facebook recruitment campaigns. We found no concern about the interpretation of the colour diagram, and the colour diagrams were interpreted as intended.

Comment #1.8
*I don’t know how to interpret attribute 6. Let’s say I have a trolley with goods costing £100 in today’s money. In a year’s time, the cost of the food will have changed relative to my salary. Perhaps my salary went down, or perhaps the food now costs £150. *But I am still able to buy it.* So the answer logically should be 100%, unless I’m down to my last shilling. I don’t think this works as a measure of loss of purchasing power.*

Reply #1.8

This attribute aims to capture standard of living. This attribute aims to capture any changes to standard of living due to changes in price and/or income (from increased taxes and/or changes in employment). For example, you may be still able to buy what is in the trolley, but this does not mean that you will have the same purchasing capacity that you used to have before the pandemic. If what costs £100 today could cost you £200 in a year (due to a price increase) or be a much larger proportion of your salary (due to a wage decline), this would mean that you will have a loss of purchasing power. Alternatively, if what costs £100 to buy 200 rolls of toilet paper in today’s money would allow you to buy only 100 rolls of toilet paper in one year, then this would mean a reduction in 100% of the trolley.

We appreciate this is a difficult attribute to frame. We have thus tested this extensively using TA interviews. As part of the survey design iterations from these interviews, we chose to describe the attribute in the survey as follows:

“As businesses close and public services adjust to the lockdown measures, the way you spend money may have changed. The lockdown measures may affect how much your household can afford to buy in two ways: you may earn less money because of wage cuts or business closures and the price of things you want to buy might have gone up. It is expected that many people will be able to afford fewer things than they could at the start of the year. The shopping trolley below shows how much **you can afford to buy** in one year compared to what you could afford to buy before the pandemic. This includes your usual shopping, housing costs and bills.”

Subsequent TA interviews indicated that respondents understood this attribute as intended.

Comment #1.9
F1 White text on light yellow background does not work well.

Consider tabbing the levels, so the attributes in bold and the levels in normal face are both left flush. I think that may make for easier reading.

Reply #1.9

Thanks for this; we have tested the white text on yellow background and found no concern on the format**.** We have tested this using TA interviews and across a range of devices. We have also included the gauge to show each type of lockdown restriction.

Comment #1.10
OSM1

I didn't "get" the linkage between the twitter analysis and the development of the attributes and levels. In the word clouds I see meaningless fragments of ideas like 'people' and 'government'. What do these contribute to the work?

Reply #1.10

The referee is correct to raise the issue of the twitter analysis and how this links to the attributes and levels. We have now added more description of how this analysis helped the development of the attributes and levels. We also drop reference to word clouds; this is likely misleading (we focused on the sentiment analysis only). The online supplementary material now reads (OSM-1):

Attributes were selected based on policy discussions and government guidance observed globally. To look more into the chosen attributes, we conducted a social media analysis. We extracted attributes-related tweets between 15th February 2020, and 19th May 2020, using the Twitter standard search application programming interface (API) consisting of a set of predefined expressions (see below), which are the most widely used news media terms relating to the novel coronavirus (COVID-19). 15,000 tweets with expressions related to the attributes were extracted for our social media analysis. Only English language tweets were extracted. We analysed the extracted tweets using **sentiment analysis,** which is a standard procedure in text mining literature.[1,2,3] **Sentiment analysis[2]:** this is the process of computationally identifying and categorising opinions expressed in a piece of text, especially to determine whether the tweeter's attitude towards a particular topic, product, etc. is positive, negative, or neutral. The key objective of sentiment analysis is to gauge opinions, identify hidden sentiments and finally to classify their polarity into positive, negative or neutral. [4]

[2] Sentiment Scale: 0-Neutral >0-Positive <0-Negative

Although we selected attributes based on policy discussions and government guidance observed globally and previous DCE studies on preferences to control emerging infectious diseases, we used the twitter analysis to gain insights into how the selected attributes are discussed on social media. We used the sentiment analysis to categorise opinions in the text related to our attributes. We identified sentiments that people have when talking about the attributes of interest. For example, do people use the term “excess death” in their tweets? What sentiments do people have when talking about excess death? Should we use excess death or number of deaths as an attribute? Are people more concerned about excess death or the number of deaths? The sentiment scores helped us to gain insights into these questions. They allowed us to identify what attitudes or sentiments people have when communicating about the attributes we selected (lockdown restrictions, lockdown duration, number of infections, excess death, hospital capacity, income loss and job loss) on social media. An analysis of the ‘excess death’ tweets, for instance, displayed 81% of negative sentiment, 4% of positive sentiment 15% of neutral sentiment. As the ‘excess death’ attribute displayed 81% negative sentiment, we were careful in the framing of the excess death attribute levels. Initially, we tested the presentation of excess death attribute in absolute numbers, but this inflated its importance relative to other attributes. Therefore, we changed the presentation of the excess death as fractions of 10,000.

Comment #1.11

The bar charts are quite amateur looking. R syntax in the axis, stretched and squashed images of different sizes, unlabelled tick marks… I can’t interpret the x axis at all.

The entire OSM1 seems like a dump for unfinished ideas.

Reply #1.11

Thanks for spotting this problem, we have improved the presentation of the bar charts and made the x axis easier to read. We have changed the x-axis label into ‘**Sentiment score**’ across all the histogram plots and added the following in **OSM-1**:

The x-axis shows the sentiment score as a negative and positive integer or zero. The sentiment score is the sum of sentiment values assigned to parts of the sentence (or textual field) and can be less than -1 or more than 1, as shown in the histogram plots. A positive score represents positive or good sentiments associated with a tweet. In contrast, a negative score represents negative or bad sentiments associated with that tweet. A score of zero indicates neutral sentiment. The more negative the score, the more negative the sentiments of the person tweeting and vice-versa.

We hope that our edits to the OSM-1, including dropping the word cloud, better explanation of the sentiment analysis, improved presentation of the bar charts and better integration of the Figures and Tables to the text, integrate it better to the protocol.

Reply to Reviewer 2

In this study, the authors provided the protocol of a discrete choice experiment to understanding public preferences and trade-offs for government responses during a pandemic. The research is important, timely and has been adequately described in this protocol. The type and number of attributes seem also

adequate. The research team has further sufficient experience to adequately perform the study. Some suggestions here below to strengthen this article:

Comment #2.1
In the development of attributes and levels, it is unclear how and when the literature search was performed. More details on the search conducted (including platform used), time period and inclusion criteria would be needed, especially as studies on this topic are quickly becoming available.

Reply #2.1

We thank the reviewer for drawing our attention to this; we have now included the following as a footnote in the **Development of attributes and levels for the DCE** section (see below) (**Page 5)**:

From May-June 2020, we searched Google (Scholar), using the following terms and combinations of them: 'covid-19', 'coronavirus, 'discrete choice experiment', 'pandemic' and 'infectious disease'.

Development of attributes and levels for the DCE

The first stage of a DCE defines the attributes and levels. Attributes describe different lockdown scenarios (Table 1) and are based on: (i) current and possible future lockdown measures from policy documents e.g. Scientific Advisory Group for Emergencies (SAGE) guidance, government guidance and interventions that have been observed globally in response to the COVID-19 pandemic (see Table 2); (ii) current literature[3] on preferences for lockdown measures;[22,23] and (iii) a social media analysis (see Online Supplementary Material (OSM-1)).

Comment #2.2
Could the authors provide more information on their experimental design and which priors/direction they will use for which parameters?

Reply #2.2

Thanks for the suggestion; we now added more information on the experimental design. The **Experimental design and construction of choice sets** section, on **Page 7,** reads as follows:

[3] From May-June 2020, we searched Google (Scholar), using the following terms and combinations of them: 'covid-19', 'coronavirus, 'discrete choice experiment', 'pandemic' and 'infectious disease'.

These attributes and levels are combined to create **lockdown scenarios** and paired into choice sets of two scenarios. In each choice task, respondents will be asked to choose their preferred lockdown scenario. We used NGENE software (ChoiceMetrics) to generate a 24 choice tasks D-efficient design with non-informative (null) priors and allowing estimation of non-linear effects of attributes.[36] Respondents will be allocated to one of three blocks, so they will neither all face the same choice tasks, nor in the same order. The design was based on the main effects only (i.e. without interactions). The 24 choice sets were blocked into three sets of eight choice tasks to reduce respondents' burden. The order of choice tasks within each block will be randomised. Scenario attributes will be presented using visual aids to ease comprehension (Table 1). Figure 2 shows an example choice task.

Comment #2.3
The authors mentioned that they "will ask respondents how likely they are to comply with the chosen scenario". Could the authors explain how these data will be analysed?

Reply #2.3

We will explore how compliance is related to factors such as moral attitudes (based on the MFQ-20), socio-economics characteristics, experiences with COVID-19 and views on government handling.

We have added the following in the **Questionnaire design** section (**Page 7):**

The compliance information will be used to understand whether people's moral attitudes affect compliance with the chosen scenario. There is some evidence of a positive correlation between high scores in the binding moral foundations and an intent to defy social distancing rules.[37] Further, the compliance data will be used to explore whether respondents who are more likely to comply with the selected scenario have a specific pattern of preferences. We will also explore the impact of socio-economic characteristics, experiences with COVID-19 and views on government handling on compliance.

Comment #2.4
For statistical analysis, it would be nice to know whether some variables will be included as categorical or continuous variable (or maybe both will be tested by the authors?).

Reply #2.4

Thanks for raising this; depending on the nature of the variables, we will code and test both categorical and continuous variables. Whilst some variables such as gender, ethnicity, country of residence, education, and income will be included as categorical variables, others such as age, household size, etc. will be included as continuous or categorical.

We now have included more information on how the variables of interest will be coded. We have now edited the text in the **Data analysis,** on **Page 10,** and the last paragraph reads as:

We will explore observed heterogeneity according to moral attitudes (using MFQ 20) and socio-economic characteristics e.g. age, gender, health status, economic insecurity, country of residence and experience with the COVID-19 pandemic. Whilst some covariates such as sex, ethnicity, country of residence, education, and income will be included as categorical variables, others such as age, household size, etc. will be included either as continuous or categorical. We will use mixed logit (MXL) models to test for unobserved preference heterogeneity, treating responses as a function of choice alternatives and individual characteristics.

Comment #2.5

It is a little bit unclear how the socio-economic characteristics will be analysed/tested in the model. Further explanation could be interesting. A latent class model could also provide relevant information and potentially be (at least) tested.

Reply #2.5

Thanks for this; the last paragraph in the **Data analysis,** on **Page 10,** is broken down into two paragraphs, and we have added the following to make the text clearer and to explain how the socio-economic variables will be analysed.

The socio-economic characteristics will be analysed by interacting them with the attributes in the multinomial logit and mixed logit model specifications. We will start from the multinomial logit model, interacting socio-economic variables with the attributes to account for observed preference heterogeneity. We will then interact the socio-economic variables with the mean of the random parameters in the mixed logit model framework to account for both observed and unobserved preference heterogeneity. We will also test latent class models, using socio-economic characteristics as covariates of class membership. Given we cannot observe individuals' moral attitudes, but rather indicators of moral attitudes, we will treat them as latent variables. To explore differences in preferences between the five dimensions of moral values, we will use a hybrid choice model with each of the parameters interacting with each of the dimensions in turn. Choice of the final parsimonious model will be determined using measures of goodness of fit (e.g. log-likelihood, McFadden's R^2, Ben-Akiva-Lerman R^2, the Akaike and Bayesian Information Criteria).

Additional References

1. Qiu, Y., Chen, X., & Shi, W. (2020). Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *Journal of Population Economics*, 33(4). https://doi.org/10.1007/s00148-020-00778-2
2. Nguyen, N. P. T., Hoang, T. D., Tran, V. T., Vu, C. T., Fodjo, J. N. S., Colebunders, R., Dunne, M. P., & van Vo, T. (2020). Preventive behavior of Vietnamese people in response to the COVID-19 pandemic. *PLoS ONE*, *15*(9 September). https://doi.org/10.1371/journal.pone.0238830

3. AlTakarli, N. S. (2020). China's Response to the COVID-19 Outbreak: A Model for Epidemic Preparedness and Management. *Dubai Medical Journal*, *3*(2). https://doi.org/10.1159/000508448
4. https://www.gavi.org/vaccineswork/here-are-four-ways-vietnam-has-managed-control-covid-19 (accessed October 2020).
5. https://www.nhsperforms.scot/hospital-data/indicator-hospital?hospitalid=3&indicatorid=28 (accessed October 2020).
6. https://breastcancernow.org/about-us/media/press-releases/almost-one-million-women-in-uk-miss-vital-breast-screening-due-covid-19 (accessed October 2020).
7. Graham, A., Cullen, F. T., Pickett, J. T., Jonson, C. L., Haner, M., & Sloan, M. M. (2020). Faith in Trump, Moral Foundations, and Social Distancing Defiance during the Coronavirus Pandemic. *Socius*, *6*. https://doi.org/10.1177/2378023120956815

VERSION 2 – REVIEW

REVIEWER	Alex Cook National University of Singapore
REVIEW RETURNED	24-Oct-2020

GENERAL COMMENTS	Thanks for considering my previous comments. One minor point: in the previous review I questioned the inclusion of a lower case i in a range (you had written j=1,...,i,...,J). You have replaced this with an upper case I (i.e. j=1,...,I,...,J). My point was intended to be that including any variable in the range didn't seem to contribute anything as the meaning is the same as j=1,...,J, which is the notation you're using for other sets. There's nothing wrong with including an i or an I in the range but it doesn't do anything so you might want to remove it. Good luck with the study.

REVIEWER	Mickael Hiligsmann Maastricht University; NL
REVIEW RETURNED	27-Oct-2020

GENERAL COMMENTS	The authors have satisfactory responded to my previous comments. Good luck with your study!