Response to the Reviews (Round 2): "How to reduce Item Nonresponse in Face-to-Face Surveys? A Review and Evidence from the European Social Survey"

January 30, 2024

Dear editors and reviewers of MDA,

Thank you for your reviews. I am very happy that the improvements are acknowledged, I myself think that the paper massively improved in the first round of reviews. I therefore get back to your minor requests as soon as possible. I again structured my response into sections.

For clarity, I will refer to the reviewer who provided the annotated PDF as reviewer 1 (R1), the other is denoted reviewer 2 (R2) respectively. As the reviews are based on the last version, page numbers will usually be from the old document ("OD"). If I specifically refer to the revised document, I add "RD".

1 Language

Both reviewers rightfully pointed to some ambiguous and lengthy sentences. R2 generally found parts throughout the text confusing. I therefore had a read myself again and improved language throughout. I will report only on major changes and the reviewer comments here though. Most of the time, I will just present the updates made in the RD.

- R2 1a: "I will therefore not distinguish between refusal and DK"

- R2 1b: added "than closed single choice items"

- R2 1c and R1 p.9: "How question design affects levels of item nonresponse is very well understood and differences between questions con-

stitute the largest part of the variance in item nonresponse (Olson, Smyth and Ganshert, 2019)."

- R2 1d: "To reduce task difficulty for members of language minorities, the questionnaire can be translated so that respondents can take the interview in the language they are most proficient in." The source is given in the following sentence: Behr and Shishido (2016)

- R1 p.12: it is indeed a personal though: "I could imagine that some of them may be mistrustful towards interviewers due to racist experiences and a fear of discrimination."

- R1 p. 16: "Based on the theory outlined above, I do not expect that the inclusion of any of these control variables or the other strategies is likely to distort the effect of another variable of interest."

- R1 p. 26: R1 pointed out that homophily is incorrect as a term here as it refers to observed network patterns, not preferences. I changed it to "strategies based on the idea that respondents trust socially similar interviewers".

2 Nonresponse and Editing

R2 Comment 2: My previous comments about item nonresponse with the editing stage has not been address. The paper still asserts, "When editing an answer, respondents may have concerns about their privacy." on p.6. Respondent may have privacy concerns when hearing the questions, far before reaching the editing stage.

I think that this is a simple misunderstanding. As mentioned on p.4, respondents can jump between stages in the survey response process. The stages do not need to be in order, although they often are. In such a case, it is indeed likely that respondents jump from comprehension to editing. I added two sentences to the paragraph to make this clear:

Respondents will likely have such privacy concerns immediately when they hear a sensitive question and jump from the comprehension stage to the editing stage in the survey response process. They probably refuse to answer before an honest answer has been formed.

3   Formula

R2 Comment 3: I don't find the formula for P_{INR} necessary for introducing the four concepts that are the core of this paper. It is unfortunately distracting. Further, without a quantification of difficulty/(ability*motivation) and privacy, you certainly cannot solve the max function involving these two terms.

The associate editor highlighted these concerns as follows:

"Comment 3 about the formula needs some attention, in the answers accompanying the revised manuscript, the author states, that the formula is not a n actual statistical formula but a help in summarizing and defining interrelationships between concept. This should be very clearly stated in the text! And perhaps the word formula should be omitted. Perhaps something like 'the interrelationship between the concepts can be formalized as...' or something similar in the own words of author. Now it still looks like an actual statistical formula."

I disagree with R2 that the formula is unnecessary and distracting. If I would just state the concepts as such, their nonlinear effects and interdependence cannot be conveyed. As this is the primary function of the formula, and not to actually provide a way to predict nonresponse, the concepts need no proper way of measurement as well.

I hope that in the revised version, this intention is now coming across even more clearly. As recommended by the associate editor, I exchanged the word "formula" for "theoretical model" to avoid confusion with a statistical formula and I introduce the equation as a formalisation and summary. The last paragraph of that section also highlights that the equation is not intended to give predictions and that the concepts need not to be measured for the purpose of it.

4   Aim and Causality

R2 Comment 6: The authors mention that they will not speaking causal effect on p. 17. Then, what is this paper for? The paper proposes the four concepts as "sources" of item non-response, which implicitly suggests a causal mechanism.

The empirical part of the paper does not set out to test whether the four concepts have a causal relationship to item nonresponse. The literature review shows overwhelming empirical evidence that they do. My empirical contribution aims to derive possible strategies to reduce item nonresponse and test whether they indeed are associated with a change in item nonresponse. So, R2 is right that the theoretical model suggests causal mechanisms for these strategies. That is the whole reason why I believe these strategies are worth investigating.

However, my empirical analysis needs to fulfill the conditional independence assumption to estimate an unbiased point estimate for the causal effect of a strategy, which is a very strong assumption. I expanded the last paragraph of the section on controls, where I make the claim R2 mentions, to better explain why I do not call my estimates causal but still believe them to be informative. I also made an effort to communicate my aim of the empirical analysis more clearly in the first paragraph of the section "Strategies to be tested".

5 Standard Errors

> Comment 7 (R2): " "Standard errors are clustered by the interviewer" on p. 18. Standard errors cannot be clustered. Please clarify. "

I agree with you that the name "clustered standard errors" is probably not the best, standard errors cannot be literally clustered. However, I did not coin the term and it is widely used in econometrics by now. I also provided a specific reference for this recommendation and I have outlined the reason why the standard errors are "clustered" by interviewer in the supplementary material (see p. 10) before submitting the OD.

As it is an established term and quite well known in econometrics, I am hesitant to changing the wording, as the more appropriate name "Liang-Zeger standard errors" is much less common.

6 Table 1

> R2 Comment 4: "Expectation" column in Table 1. Expectation for what? Expectation in nonresponse or nonresponse reduction?

By expectation I refer to the expected sign of the coefficient if the strategy works. R1 also requested an explanation of table 1 in the text. I added a paragraph before the table that explains it:

> Table 1 summarises the selected strategies that I am going to test in my empirical analysis. The second column shows which concepts play a role in the hypothesized mechanism linking the respective strategy to item nonresponse. The third column gives the expected direction of the relationship between strategy and item nonresponse, e.g. the longer the questionnaire, the more item nonresponse. These are also the expected signs of the coefficients if the strategies work as imagined.

7   The two meanings of DK

> R1 p.15: Also in your to the reviewers you state that they do not bias the results. In my opinion you should explain exactly why this is the case... Assuming they are evenly distributed and uncorrelated to any of the independent variables or covariates?

R1 is right, I assume the two mechanisms that can lead to DK, geneuine and harmful, to be uncorrelated so that my estimates are unaffected. I cannot think of a way how the possibility of genuine DK answers should bias my estimates. I have extended the respective paragraph in the subsection on dependent variables to make clear that this is an assumption on my behalf:

> Secondly, an additional mechanism that generates item non-response may increase the overall variation in the dependent variables but if it is uncorrelated to the other mechanisms, no bias in estimates is to be expected. I do not know how the possibility of genuine DK answers could interfere with the other mechanisms. I therefore assume that they are uncorrelated.

8   Controls

> R2 Comment 5: "Interferences and other people being present during the interview as well as whether respondent and interviewer have matching gender can be regarded as unrelated to any of the four theoretical concepts outlined in the model." on p.16 Why?

I am simply not aware of a mechanism potentially biasing these estimates. I changed the wording to better reflect this reasoning:

> I am not aware of any mechanisms that could lead to biased estimates for the effects of interferences and other people being present during the interview as well as whether respondent and interviewer have matching gender.