**PCI Registered Report Peer-review: Stage 2 submission**

I have undertaken the Stage 2 review of the manuscript “Does alleviating poverty increase cognitive performance? Short- and long-term evidence from a randomized controlled trial” by Szaszi et al. Overall this is an interesting and well-written manuscript which has important implications for the field; the inconclusive results will provide a more balanced literature and a point of reference for future work. However, I do have three main points for revision centering on **clarifying the “Hypotheses and Data Analysis Strategy” to avoid reader, ensuring that floor and ceiling effects are fully ruled out by clarifying them, and revising the Discussion section to ensure that the authors conclusions are justified given the evidence. I detail these as below as well as some minor revisions to improve the written style and structure further.**

**To note to the PCI Recommender,** the OSF link provided on the PCI-RR website did not successfully take me to the project page (<https://osf.io/2r8a9/?view_only=1781fb681edc4cdeb61287172cd14ba2> - I got the error “Page not found”). However, I was able to access the data through the OSF link associated with the preprint (<https://osf.io/qymaz/> via <https://psyarxiv.com/4gyzh>). This error might be because the PCI RR link is ‘view-only’, but I mention it because it might need fixing in the PCI-RR portal.

**Major points**

**My main points for revision centre on:**

1. **Clarifying the “Hypotheses and Data Analysis Strategy”**

**I am confused by the sub-sections of “Statistical framework” and “Bayes Factor Design Analysis” as they both mention information about the bayes factor cut-offs for evidence. I am not sure which sections relate to which specific analyses (i.e., in the “Statistical framework” section it states that “BF values above 10 and below 1/10 were regarded as strong evidence for the alternative and the null hypothesis, respectively” but then in the “Bayes Factor Design Analysis” section it states that “The long-term rates of correct evidence were calculated as the proportion of iterations where strong evidence (BF > 10) was found for the existence of the effect. The long-term rates of misleading evidence were computed as the proportion of iterations where the evidence strongly supported the null hypothesis (BF < 10)”. Can this be clarified? It’s also very confusing to then have another sub-section titled “calculation of BFs” with additional information below another sub-section of “specification of the models” – can you group all of the information on Bayes factors together in a more reader-friendly way? On a similar point, you continuously refer the reader to “as described below” and “as described above” which means the reader has to scroll up and down the manuscript to remind themselves: a better structure would be to avoid such language and explain this immediately.**

1. **Stage 2 Review Criteria 2A. Whether the data are able to test the authors’ proposed hypotheses (or answer the proposed research question) by passing the approved outcome-neutral criteria, such as absence of floor and ceiling effects or success of positive controls or other quality checks.** This criterion addresses whether the data quality is sufficient to be able to test the stated hypotheses, according to the pre-specified conditions in 1E.

On Page 25, you suggest that your data showed no sign of ceiling or floor effects (“Finally, the fact that the cognitive function measures were administered as part of a 90 minutes long questionnaire, could have exhausted the participants leading to floor effects. However, our data showed no sign of ceiling or floor effects”. Can you please extend the reason why so that you explicitly meet review criteria 2A. This could be, for example, by proving the range of scores for the executive functioning tasks between the two groups. This should also be clarified in Figure 1 by giving the total range that the axis can go to. By including the minimum and maximum range on this figure, you will also demonstrate visually that this was a small, non-significant effect between the two groups.

1. Stage 2 Review Criteria **2E. Whether the authors’ conclusions are justified given the evidence.** This criterion addresses whether the claims drawn by the authors in their conclusions (including in the Discussion, Abstract, and anywhere else in the paper) are warranted by the data or evidence in hand. Note that PCI RR recommendation decisions will never be based on the perceived importance, novelty, or conclusiveness of the results.

In general I think the Discussion section needs to be more specific with regards to the interpretation of the results: whilst I understand that the results are inconclusive, the language when interpreting the results is very generic and rather ambiguous at times. Here are some ways in which you might achieve this:

On Page 24 you state: “While we cannot conclude with certainty how these differences add up and interact, we can make a few observations which can put our findings into context. First, while previously published studies used pre-post designs, here the findings are based on a randomized controlled trial which in general provides a clearer and less biased estimate of the true effect size”. Here you could expand the final sentence to explain why an RTC provides a clearer and less biased estimate of the true effect size.

On Page 24 you state: “Second, although individuals participating in the study were extremely poor, they were relatively homogeneous and unusual along some of their demographics. This may have influenced the effect in some unknown way: they were all male, from Liberia, between the ages of 18 and 35, and selected to be engaged in high levels of antisocial behavior as well as poor and often homeless.” In what precise way could this affect the results? Why would this make it inconclusive? Is there any previous research you could draw upon to suggest that these demographics might matter?

On Page 24 you state: “Third, we used paper and pencil or verbal versions of three different arrow tests, two different digit span tasks and a maze task to assess changes in cognitive functioning, while previous studies predominantly used computerized forms of cognitive control and intelligence tests. We cannot be sure how the tasks and the way we administered them impacted the results.” Okay, but in what way may they have affected the results? Could pen-and-paper instruments be more noisy? Could there be experimenter error? This was something I’d mentioned in my Stage 1 review, to which you’d explained that due to the context of the study it was not feasible to collect data using computerized means; you might want to reiterate this here.

Page 24 you state: “Fourth, in the present study, participants were provided with a lump-sum cash of $200. It is an open question, how larger cash treatment or applying monthly installments instead of lump-sum money would have impacted the results. Previous results found that monthly installments vs. lump-sum money had differential effects on people’s behaviour”. In what way do these differences effect people’s behaviour? I recommend being more specific here.

Page 24, the Discussion states: “We observed a small, positive effect on executive functions both for the short (b = 0.13) and the long term (b = 0.08) toward the hypothesized direction, but the data provide inconclusive Bayesian evidence to support or reject the effectiveness of the intervention”. I would avoid mentioning the direction of the results given the inconclusiveness of the findings: you word this better in the Abstract by stating “Our main analysis revealed that cash transfers have a nonsignificant effect on cognitive performance both for the short (b = 0.13) and the long term (b = 0.08), but these observed effects are roughly four times smaller than prior non-randomized research suggested, and the evidence is inconclusive”.

1. Does the manuscript adhere to TOP guidelines?

The data and analysis code are made openly available and adhere to TOP guidelines in this respect. However, **I am** unable to open the files “STYL\_lr\_reshuffled.dta” and “STYL\_Final\_real\_data.dta” on my computer. It appears these are STATA datafiles, but I do not have access to this programme. I am also unable to open the file “analyse code.do”. Could these files be exported to an open-source programme and uploaded onto the OSF as additional files that follow the FAIR principles (<https://www.go-fair.org/fair-principles/>)? I also think it would be useful to have a data dictionary with the open data (e.g., it is difficult to know what ‘exclusion’ refers to without this being described). This is key for reproducibility.

**Other minor points:**

1. The Abstract states: “Our main analysis revealed that cash transfers have a positive, nonsignificant effect on cognitive performance both for the short (b = 0.13) and the long term (b = 0.08), but these observed effects are roughly four times smaller than prior non-randomized research suggested, and the evidence is inconclusive (BFshort-term = 1.21, BFlong-term= 0.56).” I suggest removing the term ‘positive’ so it is clear that these findings were non-significant and small. I also suggest this phrasing throughout; the reason being is that readers sometimes pull out any information that might support their own preconceptions (i.e., that there was still a positive effect, yet it was non-significant).
2. Page 3, Introduction: comma missing between cited Authors (“Mani Mullainathan”).
3. The overview on Page 11 could be written more clearly. It could simply state (suggested changes in purple font): “Furthermore, we planned to conduct two exploratory analyses: (1) a multiverse analysis to reveal the robustness and sensitivity of the results to different analytical choices (see “Robustness tests: multiverse approach”) and (2) a mediation analysis to understand the driving mechanism behind the observed effects in the primary analysis. The mediation analysis was planned for those cases where the primary analysis revealed strong support (BF >10) for the effect however we were unable to conduct this because we found no strong support for the effects in the primary analyses”.
4. Page 14 states “Finally, we standardized the executive function index to make it comparable with other results.” Can you please clarify how this was standardized (i.e., was it z-scored)?
5. On Page 15, present tense is still used when it should now be past: “Accordingly, when calculating the BF, we will use”. Please check throughout.
6. Page 16, please refer to Figure 1 in the text.
7. Page 16 states “Accordingly, we conducted multiple versions of the Intent to-treat analyses specified in the primary analysis section with 6 alternative analytical specifications (with and without control variables x 3 different priors), across 14 alternatively processed datasets (2 exclusion criteria x 7 imputation method) predicting 14 different cognitive function measures.” Can you add “as follows” to the end of this sentence to clarify to the reader that the specific information on these specifications comes next.
8. Page 17, “We repeated all the analysis with and without the control variables as specified in the primary analysis” – it would be helpful to specify these control variables in brackets rather than ‘as specified in the primary analysis’. This would aid readability/understanding. Same with “we applied no exclusion criteria on individuals.” – specify what this is briefly.
9. Page 17, remove the term ‘new’ from the following sentence as it seems like this wasn’t planned, when it was in the Stage 1 protocol: “10 new measures of cognitive function”. Perhaps ‘additional’ or ‘alternative’ would work better here.
10. Page 17, comma included where it should not be: “First, we winsorized the continuous variables at the 99th percentile while we also excluded all the individuals**,** who did not achieve at least a 80% success rate in the arrow attention test”. Please remove.
11. Page 17, typo? “explorative by nature” should be “in nature”? Again on Page 17, “We repeated all the analysis” should be “analyses”?
12. Page 18, you state “but almost all of the specifications yielded 95% confidence intervals that included zero”; can you specify the number here rather than ‘almost all’? In addition on Page 20 you state “Using the planned or small priors, most Bayes Factors are between 10 and 1/10”, again can you be more specific? Again, same wording on Page 23 which needs to be clarified.
13. Page 18 states “The effect of cash transfer becomes larger and always positive when executive functions are assessed with arrow switching accuracy, digits forward accuracy, or digit span index, but become systematically smaller and mostly negative when measured by arrow switching RT, backward digits accuracy or maze task accuracy”. Do you want to add this information to the Abstract to extend the description of the results?
14. Page 21 states “. However, similarly to the short-term results, the way the cognitive performance was assessed seem to matter. The impact of the cash transfer program was larger (mostly positive) when cognitive performance was assessed with arrow switching accuracy, digits forward accuracy, or digit span index, but was smaller (mostly negative) when measured by arrow inhibition accuracy, maze task accuracy, and maze task RT.”. Does this change any non-significant effects to significant effects? Please clarify exactly how this changes the results – it suggests that some effects are significant when cognitive performance was assessed with arrow switching accuracy, digits forward accuracy, or digit span index.
15. The figures have the wrong numbers: Figure 1 is on Page 16, so the figures on Page 23 should be Figures 2 and 3, respectively. Please correct throughout as this has created a ‘knock-on’ effect with all other figure numbers.

Signed,

Dr Charlotte Pennington