<u>Second</u> Response to Replication Report for

"Housing, Health and Happiness" (AEJ: Economic Policy, 2009)

Replication Team: M.P. Basurto, R. Burga, J.L. Flor Toro and C. Huaroto First Replication Report Date: November 2, 2014

First Response Date: December 18, 2014

Second Replication Report Date: June 22, 2015 Second Response Date: July 20, 2015[1]

By

Matias D. Cattaneo, University of Michigan Sebastian Galiani, University of Maryland Paul Gertler, UC-Berkeley

Sebastian Martinez, Inter-American Development Bank Rocio Titiunik, University of Michigan

[1] 3ie, through a miscommunication process, ended up requesting us to respond to the replication report twice, first in December 2014 and then again in July 2015. The replication team added a minor change in their replication report between our first and second response. Therefore, this document contains the same response as the one we submitted in December 2014, with the only addition of footnote 4 and a new section titled "Survey Design and Households Surveyed" that addresses, for the benefit of potential readers, a minor new additional point added by the replicators in their second replication report.

Note from the 3ie Replication Paper Series editors

3ie copy edits the papers in the series for tone. The original authors' reply alerted the editors to the possible misleading nature of a sentence in the abstract of the replication study. The editors deleted this sentence and notified the original authors of the change. However, the original authors elected not to edit their reply.

Summary

This document provides a formal response to the replication analysis conducted by Basurto, Burga, Flor Toro and Huaroto (November 2, 2014), and funded by the International Initiative for Impact Evaluation (3ie), on our article "Housing, Health and Happiness" published in the *American Economic Journal: Economic Policy* 1(1): 75-105, February 2009.

The replication analysis includes three parts: (i) a Pure Replication Analysis (PRA), (ii) a Measurement and Estimation Analysis (MEA), and (iii) a Theory of Change Analysis (TCA). In our view only the first part of the replication analysis is strictly a replication of our work,[2] while the other two are not. The second part (MEA) concerns judgment calls taken during the research project and, sometimes, requested during the peer review process that our article went through before publication. The third portion (TCA) concerns "new research" conducted by the replication team based on our original data and research design and as such it is not a replication but rather a collection of new findings that extend and build on our article.

In reading the replication report, we focused almost exclusively on its first part (PRA), which we believe might be useful insofar as it provides a double check on our data work and corresponding empirical findings.[3] Below we summarize our reading of the pure replication analysis results. We also discuss briefly the other two portions of the study, although we believe these should not be part of a so-called replication study as they are based on arbitrary and subjective decisions which may or may not be valid or justified.

In our view, the content of the MEA and TCA portions should be held to the same standards as all other research efforts, that is, a serious, objective and anonymous peer review process (like the one our original article was subjected to). For example, in the MEA part of the document, the replication team employs some methods for imputation of missing values of their choosing, which are different from the one we chose, but it is not obvious that these methods are applicable or that they reflect the state-of-the-art in the literature on missing data and imputation. A proper review process at a leading, respectable academic journal would address the latter issue, but we do not believe this should be our responsibility in our response to the replication team. As another example, in the TCA part of their document, the authors focus on different parameters of interest

---

[2] A replication analysis should not be mistaken with a replication study, i.e. a new study that assesses the effects of a prior intervention in a new population. Replication studies of relevant and internally valid research both in similar and different environments is of primary interest in social sciences.

[3] However, as it is the case for all goods produced in the economy, it is desirable to prove that the benefit of its production level exceeds the costs, not only the private but also the social costs incurred in the production of the good. Better institutional designs always facilitate a reduction in transaction costs which in turn is conducive to the production of more goods in the system, *ceteris paribus*.

(i.e., other treatment effects) employing instrumental variables and conditional techniques. Whether these approaches are indeed appropriate or justified is debatable, and again a proper peer-review process should address these and related issues. More to the point, and perhaps paradoxically, our previous working paper version included some of those estimates, but we were explicitly asked to remove them during the review process at *American Economic Journal: Economic Policy*. (See World Bank Policy Research Working Paper No. 4212, available at Social Science Research Network since August 1, 2007.)

Pure Replication Analysis

During this phase of the replication study, we were contacted many times by the replication team. We tried to be as responsive as possible, while also attempting to preserve our time and continue with our many other professional responsibilities.

As the replication report highlights (part PRA of the document), our data processing and corresponding empirical results were replicated almost completely without any complications. We are very pleased to know that the replication team reached this conclusion. It confirms that all the empirical results reported in our published article are in (almost) perfect agreement with the underlying sources used during our original research work.

As explained in the replication report, the replication team identified two very minor coding errors when constructing some variables in our final database (one placebo outcome and two independent variables). These coding errors led to slight changes in the final numerical results reported in our published paper. Correcting these coding errors does not lead to any qualitative change in the findings of our paper, and most of the actual numbers reported change very little. If anything, the results appear to be more aligned with our conclusions: for example, see the updated Table 3 where the discrepancies between control and treatment groups means are reduced.

Measurement and Estimation Analysis

The replication team decided to employ some techniques from the missing data and imputation literature to handle the data attrition present in our raw household survey data, as alternatives to the method we employed in our original research work. Whether these alternative methods are appropriate techniques to employ in this case is a judgment call, which we believe is beyond the scope of this response. For example, it is unclear whether these techniques reflect the state-of-the-art in the literature, or whether the underlying assumptions are satisfied in this particular case.

In this opportunity, the replication team found our empirical results to be “robust” to changes in the imputation and related missing data adjustment techniques. In other words, the main empirical results reported in our published paper were found to be qualitatively unchanged when some (of many possible) alternative imputation techniques were employed.

Theory of change analysis[4]

The replication team also decided to report empirical findings from two other empirical strategies based on our dataset and research design. These approaches estimate other "treatment effect" parameters---assuming certain assumptions hold. Specifically, they employ an instrumental variable (IV) strategy and a conditional strategy. Under some assumptions, the former method attempts to estimate a treatment effect on the treated parameter while the latter seeks to estimate heterogeneous treatment effects. As mentioned above, we do not view these efforts as part of a replication study, but rather as new empirical research that builds on and extends our original project. As a consequence, we believe it is beyond the scope of this response to address their findings directly. Perhaps, these empirical findings could be reported in a research paper, which could be submitted to a peer-reviewed academic journal, where the research work would be properly evaluated.

Survey design and households surveyed

After our first response to the first replication report was submitted to 3ie (dated December 18, 2014), the replication team added in their second replication report (dated June 22, 2015) a new comment. In particular, the abstract of the second replication report now includes the following sentence: "However, we were unable to fully replicate the sample section despite the use of several methods".

First of all, we note the lack of precision in the language used by the replication team: "sample section" can be interpreted in different ways. For the benefit of the readers of the replication report, we clarify here that the replicators actually refer to the survey design stage, that is, which households were interviewed during the field work conducted by the Mexican's Instituto Nacional de Salud Publica (INSP, Spanish for National Institute of Public Health). In particular, their comment does not refer to the sample employed for the empirical analysis in our paper. We were highly surprised by this added comment in their second replication report because replicating the actual survey design was not part of the replication plan. Furthermore, this survey design replication exercise is quite unusual, and infeasible given the information available to the replicators. Thus, we are not surprised by the fact that the replicators were unable to replicate the survey design because the household survey selection in our paper was done employing Census information not publicly available, and by means of a sophisticated (and computationally intensive) nearest neighbor matching technique, which was ultimately executed by researchers at the INSP. Since we informed the replicators about this, we find this added comment in the abstract of the replication report inaccurate, inappropriate and, arguably, misleading to the readers of their report, something we hope to correct with this added section to our original response note.

---

[4] Although we title this section following the terminology used by the 3ie replication report, readers should not infer from this fact that we believe the parameters or statistical analysis discussed in this section address the issue implied by the title.

In addition, it is perhaps worth mentioning here that, as any other empirical study employing household survey data, the findings may or may not represent a well-defined larger (sub-)population (external validity). Certainly, the added comment discussed here, if it were relevant, does not compromise at all the internal validity of our study, and, in fact, the exercise proposed for the replication team does not address the issue of external validity either.

Conclusion

We thank the replication team for a professional pure replication effort, which we believe was done with extreme care. We believe that the results from this effort, reported in the first portion of the replication report (PRA) enhance the credibility of the main results reported in our published paper "Housing, Health and Happiness" (AEJ: Economic Policy, 2009).

We believe pure replication efforts are important, as they provide reassurance on results published in academic journals. We also believe that subsequent research based on (and motivated by) published work could be important in advancing scientific knowledge. However, we do not regard the latter as part of a replication study, but rather as endeavors that build on and extend previous results and as such produce new academic research. By implication, we believe that the second (MEA), and especially the third part of the replication report (TCA) should be evaluated by a peer review process, and thus we did not to assess the importance or correctness of the results reported.