

It is evident that the replicators did a significant amount of research on mobile money, reviewed the paper carefully and thought about what supplemental analyses they could conduct in addition to the pure replication. These include: 1) correcting for potential differential attrition; 2) conducting heterogeneity results by age and education and the number of children under 5; and 3) changing the nutrition variable. We appreciate this list of potential analyses, and think that some of them could add value to the current research.

That being said, there are some elements of the replication proposal that suggest some misunderstandings about the paper and its findings, the data available and what can and cannot be done with the existing data. In addition, for some of these analyses, there are some concerns that the proposed analyses will yield biased estimates, either because they use endogenous controls or because they involve heterogeneous effects for variables that were not stratified at baseline, which could lead to high false discovery rates. Not recognizing this potential bias in advance may lead to statistically significant results (or results that are not statistically significant), which may be misleading. In these cases, I suggest that the replicators explain their hypotheses for these analyses and suggest how they may correct for bias or incorrect s.e.

We highlight these issues in the order that they appear in the replication proposal for their consideration.

1. *Version of the paper used for the proposal and replication.* Unfortunately, this paper had a number of missteps during the publication process. The published version of the paper did not include all revisions made to the proofs. An erratum was issued for the primary tables in the paper, but not for the appendices. The revised main tables are included in the online version of the paper, but the online appendices on the *EDCC* are incorrect. In addition, there were some small errors as compared with the online published version. These are included in a Corregendum. Therefore, before any replication occurs, it would be important for the replicators to use the correct version of the paper, and to clearly state which version they are using for their replication. This is all documented on Jenny Aker's website.
2. *Mechanisms.* The authors suggest that the paper claims that the mechanisms for the results are reduced costs, decreased timing, magnified use of mobile phones and intra-household dynamics. In the paper, the authors test for all of these mechanisms, but only state that two mechanisms are at play – reduced costs and intra-household dynamics – not the other ones. And, even then, the authors are extremely careful, and state that these mechanisms are only suggestive. On p. 3 of the published version, the authors clearly state that “In addition to the time savings, we provide suggestive evidence that the m-transfer mechanism affected intra-household decision-making.” This is again repeated on p. 22. Thus, we have tested some potential mechanisms and there are some statistically significant effects, but these are not conclusive. We wanted to clearly state this at the outset, as it seems as if it may have been misunderstood that other mechanisms are at play.

3. *Policy relevance.*

- a. The replicators state that “Aker et al (2016) provide evidence about the potential impact of using the m-transfer system to solve diet diversity” in a section on malnutrition. But in our paper, we find no impacts on nutritional status – only diet diversity – and primarily for certain food groups. We state that this could be due to our limited sample size for the nutrition data or low diet diversity – our findings are too imprecise to say whether there is an effect or not. In the table on diet diversity and nutritional status (Table 5 in the Aker version), our paper provides no evidence that the mobile money transfer system improves nutritional status, and should not be interpreted as such.
 - b. In this section, the replicators also state that “Niger can send and receive money to and from Benin and CI...It is expected that this increased the welfare of people because a money transfer through mobile money are instantaneous, secure, traceable, and dramatically cheaper in contrast with others formal and informal channels.” In fact, our analyses show that households did not use mobile money to make transfers – the coefficients were all zero. (Table 8, Panel B). And as we state in the conclusion, “In addition, while program recipient households in our study used mobile money to receive their transfer, they did not use it to receive remittances or to save, two important aspects of financial inclusion. This is potentially related to the limited m-money agent network in the country, a common issue in other West African countries.” Thus, we never state that mobile money in this context was being used for remittances. We feel that it is important to recognize that we were very careful not to make this claim in the paper or to policymakers.
4. *Attrition.* In addition to the pure replication, the replicators propose using different imputation and bounding techniques to overcome the “missing data” problem due to non-differential attrition, mainly using Lee Bounds. This is possible and feasible – and something that wasn’t done in the original paper, which recorded an attrition rate of approximately 7%. That being said, there are several things to keep in mind:
- a. The replicators refer to Gertler et al in their proposal, stating that “Gertler et al (2011) mention that a common rule of thumb for impact evaluation quality is to have less than 5%.” In particular, Gertler et al state that “Best-practice impact evaluations aim to keep nonresponse and attrition below 5 percent.” While overall rates of attrition are important, so are the rates of differential attrition across different groups, as well as the reasons for attrition. In our paper, there is not differential between the Zap, Mobile and Cash groups (Table A1), and the reasons for attrition do not appear to be driven by observable characteristics (below). That being said, there could be differences in unobservable characteristics that are driving attrition between the different groups. So, techniques that address these could be a valid check.

VARIABLES	(1) attrit10	(2) attrit10	(3) attrit10_hh	(4) attrit10_hh	(5) attrit11	(6) attrit11	(7) attrit11_hh	(8) attrit11_hh
Ittmobile	-0.01 (0.02)	-0.00 (0.02)	0.01 (0.01)	-0.00 (0.01)	-0.02 (0.02)	-0.00 (0.02)	-0.00 (0.01)	-0.00 (0.02)
Ittplacebo	-0.00 (0.02)		0.01 (0.01)		-0.02 (0.02)		-0.00 (0.01)	
Seed	-0.02 (0.02)	-0.02 (0.02)	-0.01 (0.01)	-0.01 (0.01)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)
hhage_baseline	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
livestock_baseline	0.03** (0.01)	0.03** (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)
cowpea_baseline	-0.04 (0.03)	-0.04 (0.03)	-0.02 (0.02)	-0.02 (0.02)	-0.06 (0.04)	-0.06 (0.04)	-0.03 (0.02)	-0.03 (0.02)
_Icommune_2	0.01 (0.03)	0.01 (0.03)	0.03* (0.01)	0.03* (0.01)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.02)	-0.01 (0.02)
_Icommune_3	0.04 (0.03)	0.04 (0.03)	0.04*** (0.01)	0.04*** (0.01)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.02)	-0.01 (0.02)
_Icommune_4	0.01 (0.04)	0.01 (0.04)	0.01 (0.01)	0.01 (0.01)	0.00 (0.03)	0.00 (0.03)	-0.00 (0.02)	-0.00 (0.02)
_Icommune_5	0.02 (0.04)	0.02 (0.04)	0.02 (0.01)	0.02 (0.01)	-0.01 (0.04)	-0.01 (0.04)	-0.03 (0.02)	-0.03 (0.02)
_Icommune_6	0.01 (0.03)	0.01 (0.03)	0.04*** (0.01)	0.04*** (0.01)	-0.02 (0.03)	-0.02 (0.03)	-0.01 (0.02)	-0.01 (0.02)
Ittcash		0.00 (0.02)		-0.01 (0.01)		0.02 (0.02)		0.00 (0.01)
Constant	0.04 (0.04)	0.04 (0.04)	-0.01 (0.02)	-0.00 (0.02)	0.09* (0.05)	0.07 (0.05)	0.03 (0.03)	0.03 (0.03)
Observations	1,097	1,097	1,097	1,097	1,072	1,072	1,071	1,071
R-squared	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01

- b.* The replicators propose using a missing value imputation technique and Lee bounds to deal with the missing data problem, primarily related to attrition (as some missing data are missing because they were conditional upon a previous response). Using bounding that drops observations, such as Lee bounds will reduce the sample size and reduce precision. A careful read of the Aker et al (2016) paper would suggest that: 1) only the upper bounds would be statistically significant for most of the results where there are impacts (ie, Table 4, Panel A and Table 5, Panel A), not the lower bounds; and 2) for those results that are only statistically significant at the 5 or 10 percent levels in the paper, the upper bound may also not be statistically significant, especially due to reduced precision. We

would hypothesize that only the upper bounds would be robust to Lee bounds, and only for those effects that are statistically significant at the 1 percent level.

5. *Heterogeneity results.* The authors propose testing for the heterogeneity of impact on diet diversity, focusing on age and education. While heterogeneity analyses can often yield important insights, and were not conducted by the authors, there may be some important considerations to keep in mind:
 - a. *False discoveries in testing for heterogeneous effects.* While heterogeneity analyses are common, there are also some potential drawbacks of this approach. First, as the experiment did not stratify by age or education prior to randomization (the randomization occurred at the village level), the experiment was not intended to test these heterogeneous effects. Fink et al (2014) show that testing for heterogeneous effects ex-post – ie, without stratifying prior to randomization – can lead to a large number of false discoveries; as a result, a FDR correction should be done on these analyses.
 - b. *Age and education variables.* The age and education variables being used for these tests typically have classical measurement error. Classical measurement error in the RHS variable will lead to attenuation bias for that variable, and also lead to other bias (either upward or downward) for other RHS variables. It would be helpful to take this into consideration during the interpretation of the results.
6. *Theory of Change analysis.*
 - a. The replicators propose modifying the nutrition analyses to transform the nutrition variables in different ways, mainly by converting these variables to binary variables for prevalence of severe malnutrition. This is definitely possible and feasible, and would be welcome. It should be noted that these analyses were already done by the authors (although not included in the paper), as no effects were found. It should also be noted from the outset (as is stated in the paper) that the number of observations of the nutrition data is a subset of the total experiment – about 1/3 – and so power is very limited.