Technical Proposal

3ie Replication Window 4: Financial Services for the Poor

Replication of "Cash or Condition? Evidence from a cash transfer experiment"

Maira Emy Reimão*

June 2017

Introduction

This technical proposal pertains to the replication of "Cash or Condition? Evidence from a cash transfer experiment," by Sarah Baird, Craig McIntosh, and Berk Özler and published in the *Quarterly Journal of Economics* in 2011 (vol. 4). The article is part of a larger research program co-led by the authors in the Zomba district of Malawi, and based on offering cash transfers to girls aged 13-22 – conditional or unconditional on school attendance – over a course of two years (2008-2009). Treatment assignment was done through a multi-step process: (1) enumeration areas (EAs, each covering about 250 villages) were randomly assigned into an unconditional cash transfer (UCT) arm, a conditional cash transfer (CCT) arm, or a control group; (2) within the UCT and CCT enumeration areas, villages were randomly assigned an amount for the parent cash transfer; and (3) in the UCT and CCT enumeration areas, each eligible girl was randomly assigned an individual cash transfer (N=2907).

The conditional transfers were given to girls and their parents – two randomly assigned amounts – on the months in which the girls had at least an 80% attendance rate in school. School fees were also covered in the conditional group when applicable.[1] In the unconditional group, both the girls and their parents received a randomized cash transfer, and an additional amount was given the household to cover school fees when applicable. The random portion of the transfer averaged $3 per month for the girls and $7 per month for the parents in both the UCT and CCT arms; and the additional transfer for UCT households whose girls were eligible for secondary school was roughly $7 per month (for more details, see Baird et al, 2011: footnote 20). Households in the control villages did not receive any transfers.

"Cash or Condition" investigates the effect of the CCTs and UCTs on the primary goal and related outcomes of the intervention – namely, school attendance and learning (as measured through test scores) – as well as other desirable outcomes – avoided early pregnancies and marriages. The authors find that the CCT increased school enrolment and some test scores, while the effect of the UCT on these education outcomes was not statistically significant. In contrast, the CCT had no effect on the incidence

* Post-Doctoral Fellow, Yale University and Evidence Action. maira.reimao@yale.edu.
Julian Rose (University of Goettingen) will provide research assistance for this replication.
[1] There are no school fees in Malawi for primary school students, but school fees are applied in secondary school. (Baird et al., 2011: footnote 18.)

of marriage or pregnancy by the end of the intervention (early 2010), while the UCT significantly decreased the likelihood of both outcomes (by almost half and by a quarter, respectively).

As the authors explain, these seemingly opposing results are driven by the mechanics of the transfers on school drop-outs. While the CCT successfully raised school enrollment and attendance for some girls, there is nonetheless a share of girls who did not attend school (30.8%) and therefore did not receive the cash transfer. In contrast, while the non-attendance rate is higher in the UCT (39.5%), girls in that arm who did not attend school continued to receive the transfers. For these girls, the boost in income from the transfer lowered their likelihood of early pregnancy or marriage. These results point to the existence of a potential trade-off between CCTs and UCTs: conditional cash transfers may encourage compliance with the direct or related outcomes, but fail to provide a safety net for households that are unwilling or unable to comply with the condition but may have nonetheless benefited meaningfully from an unconditional transfer.

Relevance of Original Paper

At the most basic level, “Cash or Condition” is a key piece in the literature comparing the impact of UCTs and CCTs, untangling the effect of the cash transfer and that of the conditionality. As more cash-transfer programs are contemplated and implemented by governments and development agencies each year, it has become increasingly important to understand the effect of various program features. And the recent proliferation of studies in this area – and the call for sorting through them – is reflected in the existence of a systematic review comparing UCTs to CCTs in education, and another on the dimensions of impact of cash transfers more generally.[2] The former shows that while UCTs do have a positive impact on school enrollment, CCT programs with strong monitoring and enforcement systems achieve greater impact in this area. And both reviews find limited evidence of a positive impact from cash transfers on learning/ school achievement. Nonetheless, the latter notes that cash transfers do tend to boost household total and/or food expenditures.

The relevance of “Cash or Condition” to research and policy is not limited to its findings on the impact of UCTs versus CCTs in education, however. Rather, it also fits into the broader literature and discussion in international development on holding UCTs as the benchmark or the “default” intervention to which other programs should be compared in terms of cost-effectiveness and impact.

Unlike a systematic review, which tends to focus on outcomes that are *shared* across studies, “Cash or Condition” draws attention to the fact that UCTs might not actually be an adequate benchmark for other programs, as they work in a unique way and may achieve *distinct* outcomes. As detailed in this particular paper, for instance, CCTs may have better success at raising school attendance and learning for girls, while UCTs can continue to provide transfers to more vulnerable households (i.e., those whose girls drop out of school even with a CCT program) and lower early marriage and pregnancy rates in this group. It is not surprising that a program that rewards households for a particular outcome actually gets better compliance with that outcome than one that does not, but a measure of the cost-effectiveness of

[2] See Baird et al 2014 for a review of CCTs and UCTs, and Bastagli et al. 2016 for a systematic review of cash transfer programs.

the intervention with respect to the desired outcome and in comparison to a UCT overlooks the fact that a UCT reaches and benefits non-compliers.

Given the contribution of this paper to the debate on holding UCTs as the benchmark in development programs, replicating it and ensuring that the empirical results are not only valid but also adequately interpreted is important for the research literature as well as policymaking in international development. In the following section I describe a proposal for replicating this research, starting with push-button and pure replications and following onto a measurement and estimation analysis (MEA), with particular attention to heterogeneous effects and endogeneity.[3]

Replication Proposal

Push-Button Replication

The dataset and Stata .do file for replicating "Cash or Condition" is publicly available through one of the author's website (https://sites.google.com/site/decrgberkozler/papers-by-topic). This in itself should not be overlooked, as it is indicative the authors' commitment to transparency and supporting replication and secondary use of data. In this same website, the author has included a document with notes comparing the published results to the results achieved through the shared replication .do file. The document shows that in the instances in which the two differ, the discrepancy is not meaningful (tenths or hundredths of a percentage point). The source of these differences is not explained, however; perhaps the data was further "cleaned" between the production of the tables and the submission of the public datasets, but this is not clear.[4]

As the dataset for push-button replication and .do file are publicly available, I expect this phase of the replication to be straightforward. Push-button replication and documentation will be conducted with support from a research assistant, and results from the exercise will be compared to both those in the published paper as well as those in the document provided in the author's website.

Pure Replication

The dataset for push-button replication is available through one of the author's website, as noted above. This dataset, however, appears to be limited to variables used in the publication and strictly needed for recreating the tables presented therein. But "Cash or Condition" is actually just one output from a large RCT co-led by the authors, and three rounds of data pertaining to the entire project are available through the World Bank's microdata catalog (http://microdata.worldbank.org/index.php/catalog/2339/data_dictionary). So, the pure replication will

[3] The current proposal has been revised from its original form to incorporate comments from internal and external 3ie reviewers.

[4] I have contacted the authors inquiring about this document, along with other items, but the authors have determined, understandably, that they will only focus on one replication of their work at a time. Another paper relating to this study is currently under replication through 3ie as well. I will provide an update regarding this document and the source of the discrepancies with my own push-button replication results if the former is available from the authors once the latter is complete.

use the latter, to ensure that when the three rounds of data are compiled by my research assistant and myself, and we process the data in a similar way, we achieve similar results.[5]

In this phase of the analysis, we will seek to replicate each of the tables prepared by the authors, including the "battery of robustness checks" included in the appendix.[6]

Measurement and Estimation Analysis

In the third phase of replication, I will use the datasets from the World Bank Microdata Catalog to carry out further checks and investigate the impact of some reasonable adjustments to the model. I emphasize here that this will not be a search for different effects nor for uncovering differences with the authors' narrative. Rather, I will explore questions that I expect would be salient to peer reviewers and important for policy decisions regarding CCTs and UCTs.

Under/over reporting of enrollment by school or teacher. Table III in "Cash or Condition" shows that UCTs appear to have a positive impact on school enrollment when using self-reported attendance, while CCTs do not. The authors argue, however, that self-reporting over-estimates attendance for girls in the UCT arm and instead favor the use of teacher-reported enrollment. Further detail for this preference is provided in a different paper that has come out of this RCT,[7] but it reverses the effect and reveals a positive impact from CCTs and not UCTs.

The puzzling part of this reversal, however, is that since UCT and CCT recipients are compared to the control group, this shift would require that girls in the UCT arm be more likely to over-report attendance than those in the CCT or control arms, and that control girls also over-report to a greater extent than CCT girls. It is possible, however, that *teachers* with CCT students are more likely to over-report attendance, as discussed in "Cash or Condition" footnote 19. While the authors dismiss that through initial spot-checks at the schools, I think it is also important to check whether discrepancies in enrollment rates vary significantly by school or teacher, as these would provide some evidence of issues with teacher-reported enrollment.[8] Moreover, the discrepancies may be due to how outcomes are measured, as it seems that the household survey asked about enrollment while the ledgers followed attendance. Perhaps UCT girls enroll *at higher rates* than CCT girls, but do not attend? Table IV in "Cash

---

[5] As in the previous footnote, I have requested from the authors the .do file transforming the World Bank dataset into the dataset for push-button replication, and will provide updates if/when received from the authors. This would allow me to see additional changes and assumptions made as well as observations dropped during the process, if any.

[6] A potential exception may be Table IV in the paper, for which the code is not provided for push-button replication, and which may not be replicable with the publicly available dataset, as it requires information from school ledgers and spot-checks, as discussed later in this proposal.

[7] Baird and Ozler, 2012. This paper is based on a comparison between the household survey data and data collected through spot-checks by the research team as well as school ledgers. It appears to me that the latter two are not publicly available, but I will investigate more thoroughly and contact the authors if needed.

[8] The latter likely can only be done if I have access to the data from spot-checks or school ledgers (see footnote 4 here).

or Condition?" introduces more questions, since it shows no *differentiated* over-reporting between UCT and control.

Treatment by schooling level given age. The authors explore the effect of treatment by age, and find that the CCT did not have any heterogeneous effects with respect to age, while the UCT was significantly more likely to lower the probability of pregnancy and marriage for girls 16 or older. (That the latter is only relevant for older girls is of course not surprising, since we would expect relatively low pregnancy rates across the board for girls younger than 16.) But age itself may not be a meaningful moderator for schooling outcomes in a context of high drop-outs and grade repetition. Rather, an effect may be differentiated by whether a girl is close to the correct grade for her age or not. Note that, at baseline, eligible girls in this study are aged 13-22 but still two or more years away from completing secondary school; girls 18 and older in this group are effectively in the incorrect grade for their age, as are likely several girls in the younger cohorts. In fact, in 2007 in Malawi, 69% of girls in Standard 8 were considered over-age, with 22% at least 3 years over age (World Bank, 2010).

I will consider then whether each of the treatments are more impactful for girls who have a history of weak engagement with school, compared to girls who have have mostly attended school and are close to the correct grade for their age. While the dataset likely does not have a full schooling history, so that I cannot know why each girl is behind the correct grade for her age (e.g., drop-out versus repetition), we can nonetheless distinguish between those who are close to the correct age and those who are not, as information was collected on which grade each girl would attend (if she were in school). [9]

Additional transfers for secondary school girls in the UCT arm. Another issue that needs to be considered is the difference in transfer size for girls eligible for primary school and those eligible for secondary school. As the authors explain in footnote 20, while there are no school fees for primary school in the region, secondary schools do impose fees. To account for this, the authors understandably made adjustments to the intervention for secondary schoolers – they covered the fees for those in the CCT arm and provided an additional transfer of $7 per month (the average cost of the school fee) for those in the UCT arm.[10] This raises an issue of heterogeneity that must be taken into account, not only because the effect of CCTs and UCTs may already differ by schooling level (i.e., it may be easier/harder to encourage girls to stay in secondary school, once they have already reached that level) but also because the transfer amount was actually different for households whose girls are eligible for secondary school and those who are not. In practice, for families with girls in secondary school, those in the CCT arm who

[9] Primary school in Malawi (Standard 1-8) is technically targeted at children 6-13 years of age; secondary school follows for 4 years thereafter (Form 1-4). So, if a respondent is aged 17 but still assigned to Standard 8, for example, we know that she is well behind the correct grade for her age, and likely repeated or dropped out previously. In contrast, a 13-year-old assigned to Standard 8 is in the correct grade for her age, and is unlikely to have previously repeated or dropped out. In my analysis I will use a buffer of an additional year (e.g., labeling 14-year-olds assigned to Standard 8 as close to the correct grade for age, for example) to account for differences in birthday cut-offs for enrollments and possible late enrollments in primary school. I will allow for interactions between this dummy for "close to correct grade for age" and treatment to allow for differences in impact for these two groups.

[10] I contacted the authors but have not yet learned whether this additional $7 transfer was made to the parents or the school girl.

did not drop out received an average transfer of $10 along with the coverage of the school fees, while those in the UCT arm received $17. In contrast, for girls in primary school, the average cash transfer to the family was $10 in both the CCT and the UCT arm.

Relative transfers to parents and students. As the intervention gave transfers to the treatment girls and their parents – with randomized amounts – studying the effect of this variation further may also be important. The marginal effects of the cash transfers to the girls and their parents are explored separately towards the end of the paper, and Baird et al. (2011) find that there is no marginal effect from the values of the CCT. That is, the positive effect of the intervention on schooling through the base transfers ($1 to the girl and $4 to the household) is not significantly different from the effect of any other higher transfer values. The opposite is true for the UCT, where incremental increases in the amount transferred to the parent (but not to the girl) increase school attendance and test scores and lower the likelihood of marriage. This result is somewhat surprising, and I will explore whether there is a non-linear relationship between transfer values and these effects, such as a threshold over which there is an effect. In the same vein, I will analyze whether the relative size of the transfers between the girls and their parents matter, as well as the size of the transfer relative to other girls of similar age in the same village (peers).

Spillovers within villages. Another potentially relevant design feature of the intervention is the fact that the share of eligible girls who were actually offered the transfer was randomized across villages (0, 1/3, 2/3, 1). It is very likely that the authors already explored the importance of this variation, but it is not mentioned in the paper. I believe it is important to conduct this analysis within this replication effort, even if only to document the lack of spillover effects within a village.[11]

Other moderating factors. In response to comments received, I will also explore whether there are other characteristics that moderate the impact of the intervention, such as whether the household is a farming household or the presence of younger siblings. The impact of the intervention on school attendance is likely different for households that rely more heavily on family labor, though it is unclear in which direction. On one hand, they may place a clearer value on girls' times, so that if the transfer is above that, the girls are more likely to attend school, while girls who do not attend school but do not provide family labor may be away from school for other reasons that cannot be compensated by a transfer. On the other hand, reliance on family labor may be binding, and the cash transfer may not be sufficient to purchase the needed labor from outside the home.

Switching regression for drop-outs. The key puzzle and interesting piece of this paper is the seemingly conflicting effects between CCTs and UCTs regarding school-related outcomes and marriage and pregnancy. In demonstrating that UCTs decrease marriage and pregnancy primarily through the impact of the transfers on girls not attending school, the authors compare the marriage/pregnancy rates by treatment and enrollment status (Table VIII) and run separate regressions by enrollment (Table IX). Enrollment status, however, is endogenous, particularly in Malawi, where one of the principal reasons

[11] I do not know if the public datasets include a variable indicating the spillover category of each village, as it was probably not collected through the household survey, but, if not, I will contact the authors to request this information as well.

cited by girls for dropping out of school is in fact pregnancy/marriage, particularly towards the end of primary school (World Bank, 2010). Therefore, girls who are less likely to be enrolled despite the treatment are also be less likely to have their marriage/pregnancy outcomes influenced by a cash transfer. This makes splitting the sample and running separate regressions enrolled and drop-out girls potentially problematic.

To better understand the relationship between treatment, enrollment, and marriage/pregnancy, I propose modeling this system as a switching regression with endogenous separation, which can be estimated through maximum likelihood.[12] In this model, girls can be in one of two regimes: enrolled in school or not enrolled, where enrollment is determined by their treatment status as well as other factors, including their grade at baseline. Identifying a variable that might determine enrollment but not marriage or pregnancy may be challenging,[13] but, with it, I can measure the effect of the treatments on marriage/pregnancy *within* each regime *accounting for* the endogeneity of regime selection.

Additional Thoughts

When considering the relevance of this study for the discussion on using unconditional cash transfers as the "benchmark" for other inventions, a moment of thought should be given to the potential impacts that are unique to this intervention carried out in Malawi. In particular, as noted in "Cash and Condition", footnote 24, girls in each treatment arm were aware of other treatment arms, which would not be the case if either treatment was implemented universally. This in itself may alter the behavior of either treatment groups (Hawthorne effect), or even of the non-treated girls, as they act to compensate for their lack of treatment (John Henry effect[14]). The discrepancy between teacher and self-reported school attendance, for instance, raises the presence of this possibility further, as it would suggest that UCT girls might feel pressure to compensate (of which over-reporting their attendance would be one dimension) relative to CCT girls. Importantly, neither of these effects can be tested directly in a replication study (as opposed to a reproduction of the study in the same or other site), and do not call into question the internal validity of the research. Nevertheless, these issues should be kept in mind in the broader debate on the relative effectiveness (and cost-effectiveness) of UCTs.

Timeline

I will follow the timeline stipulated by 3ie. Given that the datasets are publicly available, the push-button replication can be carried out quickly and results submitted promptly. Pure replication will likely take a few more months, and will be carried out during the summer of 2017. The remainder of the allotted time will be spent in the measurement and estimation analysis, allowing for sufficient time

[12] For further details on switching regressions, see Maddala, G.S. (1983).

[13] At this stage, I am considering that "grade at baseline" might be a plausible candidate, controlling for age. This variable likely affecting whether someone drops out (e.g., a girl might be more likely to drop-out if at baseline she had just completed primary school, since we know a lot of drop-outs happen in the transition from primary to secondary school) but not whether one gets pregnant or marries early.

[14] A thank you to Julian Rose for pointing me to the observation and name of this latter effect.

before final submission for comments from reviewers and the original authors of the study. A final report will be submitted within 12 months of the start of this window.

Familiarity with Datasets

As required by the rules of this replication window, please note that the current proposal was prepared once I had found and downloaded the publicly available datasets in one of the author's websites as well as the larger dataset in the World Bank's Microdata Catalog. I have seen each of the datasets and noticed that the World Bank's dataset – available online for each of the three survey rounds separately – pertains to the entire project while the one in the author's website is limited to the data necessary for push-button replication of the present paper. I have not, however, run any code to analyze any of these datasets, including gathering descriptive statistics or studying the additional variables that may be available. It is possible, therefore, that I may not be able to carry out some of the activities proposed here, if it turns out that certain data/ variables are not available. That said, I have read several of the authors' other publications related to the study, and these have also provided me with information on the contents and likely replication potential of this dataset.[15]

The replication will be conducted using Stata.

[15] An earlier version of this proposal, for example, suggested considering whether some of the school drop-outs in the study were not drop-outs, but in fact girls completing school. I learned, however, from a different paper (Baird et al. 2012) that the girls chosen for the study were limited to those with at least 2 years of education remaining, to ensure that they could benefit from both years of the study.

References and Literature Reviewed

Baird, S., E. Chirwa, C. McIntosh, B. Özler. 2009. “Research Proposal: Unpacking the Impacts of a Randomized CCT program in sub-Saharan Africa.” http://microdata.worldbank.org/index.php/catalog/2339/download/34881

Baird, S., F. H. G. Ferreira, B. Özler, and M. Woolcock. 2014. “Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes”, *Journal of Development Effectiveness* 6(1):1-43.

Baird, S., R. S. Garfein, C. T. McIntosh, and B. Özler. 2012. “Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomized trial”, *Lancet* 379:1320-9.

Baird, S., C. McIntosh, and B. Özler. 2011. “Cash or Condition? Evidence from a cash transfer experiment”, *The Quarterly Journal of Economics*, 126(4):1709-53.

Baird, S. and B. Özler. 2012. “Examining the reliability of self-reported data on school participation”, *Journal of Development Economics* 98:89-93.

Bastagli, F., J. Hagen-Zanker, L. Harman, V. Barca, G. Sturge, T. Schmidt, and L. Pallerano. 2016. “Cash transfers: what does the evidence say? A rigorous review of progamme impact and of the role of design and implementation features.” London: Oxford Development Institute.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Economics*. New York: Cambridge University Press.

The World Bank. 2010. “The Education System in Malawi: Country Status Report.” Washington, DC: The World Bank, UNESCO, Pole de Dakar, Education for All, and gtz.