DECISION MEMORANDUM

TO: Gates Foundation

FROM: Emma Stiefel and Arham Hameed

SUBJECT: Comparing Conditional and Unconditional Transfer Programs

SUMMARY:

There is increasing evidence that conditional cash transfer (CCT) programs can have large impacts on school enrollment, including in very poor countries. To further understand this impact and advise you on the implementation of a cash transfer program in sub-Saharan Africa, we examined the paper **Cash or Conditions? Evidence From a Cash Transfer Experiment** (Baird, McIntosh, and Ozler, 2010). We partially replicated Baird et al.'s results and validated their analysis, which concluded that both CCT and unconditional cash transfers (UCT) can be valuable, but for different people. We then extended their results to investigate the possibility of building a model your organization could use to determine which type of support to give a girl.

PROBLEM AND OBJECTIVES:

The Gates Foundation is implementing a new cash transfer program for schoolgirls throughout sub-Saharan Africa and wants it to have a maximum impact on the schooling outcomes of participants. Essentially, employees must decide between whether to give participants cash transfers conditional on them attending school or to give them money regardless of their attendance. Baird et al.'s study used a randomized experiment to examine similar target participants in the sub-Saharan country Malawi and found that neither CCT and

UCT is absolutely better than the other. Our objective, therefore, is to explain their study results and advise you in the implementation of the cash transfer program in deciding between which participant should get CCT or UCT.

EXPERIMENT DESIGN:

Baird et al. randomly assigned 176 enumeration areas (EAs) in a region of Malawi to treatment or control status. The 88 treatment EAs was then randomly assigned to either receiving CCT or UCT interventions, while the remaining control EAs received no support. Participants within treatment EAs were equally likely to take tests used to assess schooling outcomes or not be assigned to treatment. This type of experiment assessing the marginal contribution of the conditionality in a cash transfer program – i.e. a randomized controlled trial with one treatment arm receiving conditional cash transfers, another receiving unconditional transfer, and a control group receiving no transfers – has not previously been conducted anywhere.

The overall data was collected in three rounds. Round 1 is the baseline data that was collected before the invitations to participate in the program were sent to treatment households. It was possible, therefore, that some participants would not enroll in either the CCT or UCT program despite being invited and even that the field agents administering the program enrolled some who were not assigned to treatment. Baird et al. do not explicitly address these compliance rates and seem to suggest that all who were invited to participate in the study did. Because families would lose almost nothing from participating in either CCT or UCT, perfect or at least very high compliance seems feasible, so we do not think that this potential problem compromised Baird et al.'s conclusions.

Round 2 and 3 data were collected approximately at a 12 months period after each other. The two rounds' data was collected after the completion of the 1st year intervention and 2nd year intervention respectively. Overall 2,907 girls were surveyed at round 1 and 2,089 girls participated in all three of the surveys, for an overall tracking rate of over 90%, further increasing our confidence in their results.

Baird et al. measured variables that were strongly predictive of schooling outcomes. The balance of the experiment on these and other demographic characteristics was tested using the baseline data for the sample used in the analysis. The experiment appears to be well balanced between the treatment and the control over a broad range of outcomes; only age and highest grade attended, two highly correlated variables, differ significantly between the two groups. The similarity between the control, CCT, and UCT groups at baseline increases our confidence that the randomization of this experiment was successful and there was no selection bias or significant confounding variable distorting the results.

Baird et al. thoroughly address other potential problems with their study. Once the requirement for an experiment's causal validity, for example, is that participants don't affect each other's outcomes. That condition could have been violated in this study because the EAs were close to each other and participants with different treatments would likely interact. It could be possible, for example, that UCT girls would go to school more often not because of UCT itself but because they are influenced by their CCT peers who try to meet the attendance requirement. Baird et al. assessed the extent of this potential problem by evaluating outcomes of all the girls at each Cash Transfer Location, buildings where girls from both CCT and UCT EAs would go to receive payments, and found that the outcomes of UCT girls did not increase if they lived near CCT girls, successfully mitigating the potential concern.

After seeing that Baird et al. sufficiently addressed this and other concerns we had with their experiment design, we were convinced that their procedure was sufficiently random and unbiased. We did not need to perform procedures common in similar studies that use observational data, such as analyzing how sensitive results are to an unobserved variable or pairing treatment and control observations to make sure the groups are comparable, because the treatment assignment was indeed random, resulting in groups that could be validly compared.

REPLICATION:

Baird et al. presented the results of the program impact on the Educational Achievement on the two treated arms in Table VI. Because test scores are such an important outcome (school attendance means little, after all, if students aren't actually learning while they're there) we decided to replicate this table.

To get the impact estimates listed in the table, Baird et al. used a model to determine the increase of switching a girl from no intervention to CCT or UCT while controlling for baseline variables that are strongly predictive of schooling outcomes. Because Baird et al. did not explicitly include the specific functions and, in some cases, variable names they used in their paper, we could not exactly replicate their model or results. Our replication, however, was still very close to their original results, actually increasing our confidence that the conclusions they draw are valid and not dependent on slight assumptions they made that we could not copy.

Table VI: Program Impacts on Educational Achievement

	Dependent Variable:			
	English test score (standardized)	TIMMS math score (standardized)	Non-TIMMS math score (standardized)	Cognitive test score (standardized)
	(1)	(3)	(2)	(4)
Conditional treatment	0.140***	0.120*	0.086	0.174***
	(0.054)	(0.067)	(0.057)	(0.048)
Unconditional treatment	-0.030	0.006	0.063	0.136
	(0.084)	(0.098)	(0.087)	(0.119)
Observations	2,057	2,057	2,057	2,057
Prob > F(Conditional=Unconditional)	0.069	0.276	0.797	0.756

Notes: The cognitive test score is based on Raven's Colored Progressive Matrices. Math and English reading comprehension tests were developed based on the Malawian school curricula. Five questions (four from the Fourth Grade test and one from the Eight Grade test) from Trends in Mathematics and Science Study (TIMMS) 2007, which is a cycle of internationally comparative assessments in mathematics and science carried out at the fourth and eighth grades every four years, were added to the Math test. All test scores have been standardized to have a mean of zero and a standard deviation one in the control group. Regressions are OLS models using Round 3 data with robust standard errors clustered at the EA level. All regressions are weighted to make the results representative of the target population in study EAs. Baseline values of the following variables are included as controls in the regression analysis: age dummies, strata dummies, household asset index, highest grade attended, an indicator for ever had sex, and whether the respondent participated in the pilot phase of the development of testing instruments. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Replication of Table VI: Program Impacts On Educational Achievement

	English Test Score	**TIMMS Maths Score**	**Non-TIMMS Maths Score**	**Cognitive Test Score**
Conditional Treatment	0.143 **	0.113	0.086 *	0.197 ***
	(0.050)	(0.060)	(0.053)	(0.057)
Unconditional Treatment	-0.020	0.004	0.071	0.162
	(0.073)	(0.087)	(0.080)	(0.090)
Observations	2,057	2,057	2,057	2,057

EXTENSION:

Baird et al. found that CCTs increase schooling outcomes, but UCTs are better at delaying marriage and pregnancy. They speculate that this is because girls who are already at risk of dropping out and starting a family can't fulfill the attendance requirements to receive CCT, and so end up with no additional income and the same likelihood of being married as a girl

from the control group, but can afford to delay marriage if they receive unconditional support. Based off this analysis, they advise that CCT is best for younger girls who are less likely to drop out to start a family but should be given UCT when they turn 16, the age where test score differences between the two groups taper off but UCT starts to become much better at preventing marriage.

Age isn't the only predictor of marriage and pregnancy, however. We found that, among girls surveyed in round 3 of the study who are less than 16 years old, 7% were married, 9% had been pregnant, and 5% were both married and pregnant. These young mothers and wives may be a tiny minority of girls, but they are perhaps the most vulnerable members of the study population. We know that you value helping all people in need, so we wanted to see if it is possible to build a model for identifying which girls would do better with a UCT that improves on Baird et al.'s proposed 16-year-old cut off.

To do this, we decided to build a relatively primitive test model that would predict whether a girl with a certain set of baseline characteristics would do better with a CCT or UCT. We didn't try to make our model perfectly predictive; we just wanted to see if we could come up with something that is more accurate than using age alone.

We began building our model by pairing up girls from the CCT and UCT groups who were similar at the beginning of the study, as measured by the main variables Baird et al. used in their analysis: household asset index, age, highest grade attended, whether their parents were alive, if they were a virgin, and the geographic strata they lived in. Because we already know how well each girl did in school and whether she was married or impregnated, we can see which of these similar girls was more successful overall: the one that received CCT, or the one that received UCT. In order to do this, we calculated an "overall success

score" for each girl that combined the marriage/pregnancy and schooling outcomes into a single value.

For each of these similar pairs we created a single hypothetical girl with characteristics that are the average of the original two and paired this set of characteristics with the difference in overall success between the original pair. We used part of this newly constructed data to teach our model how to predict whether a certain set of baseline characteristics suggest that a girl would do better with UCT or CCT. When we tested our model on the rest of the data, we found that it was approximately 58% accurate. The simple 16-year-old cutoff was less accurate, correctly predicting if a UCT or CCT would be better only 46% of the time.

Our model still needs improvement, but we have shown that it is possible to improve on Baird et al.'s 16-year-old cutoff guideline. We recommend that further work be done to develop this predictive model (such as developing a machine learning model) so that field agents can use it to determine what type of aid a girl should receive.

CONCLUSION:

We examined Baird et al.'s experimental study of the impact of CCT and UCT on schooling and delayed marriage/pregnancy outcomes. The experiment overall is unbiased and Gates Foundation should certainly implement the new cash transfer program for school girls. The results showed that CCT and UCT arms were more likely to be enrolled in school compared to the control group. Furthermore, our extended model makes it more likely to predict correctly whether a girl should get a UCT or a CCT. This will help the Gates Foundation to target individuals at risk and provide the appropriate treatment at an individual level that suits them the most.

Appendix:

Code PDF:

https://drive.google.com/file/d/1KK1C15Xxg-prthjkxTKUz5sGkAlkWLqD/view?usp=sharing

Code .R:

https://drive.google.com/file/d/1OP9SaBbEbfBF1mUwu80LAau1C88sxKLv/view?usp=sharing

---

Emma's contribution: I wrote the code for our extension model, almost all of the code for the replication model, wrote the extension and replication section of the paper, and revised the rest of the paper.

Arham's contribution: Worked on the code to find the sample size they used, worked on the Experimental Design, Problems, Objective, and Summary of the Decision Memo.