

Title Page

Conditional, Unconditional and Everything in Between: A Systematic Review of the Effects of Cash Transfer Programs on Schooling Outcomes

Sarah Baird: George Washington University/University of Otago. PO Box 56 Dunedin, New Zealand. Email: sarah.baird@otago.ac.nz. Phone: +64 3 479 3042. (Corresponding Author)

Francisco H. G. Ferreira: The World Bank. 1818 H Street, NW, Washington, DC 20433. Email: fferreira@worldbank.org. Phone: 202-473-4382.

Berk Özler: University of Otago/The World Bank. PO Box 56 Dunedin, New Zealand. Email: berk.ozler@otago.ac.nz. Phone: +64 3 479 8643.

Michael Woolcock: The World Bank. 1818 H Street, NW, Washington, DC 20433. Email: mwoolcock@worldbank.org. Phone: 202-473-9258.

Funding: This work was supported by the Australian Agency for International Development (AusAID) under Grant Agreement Number 59616. The views expressed in the publication are those of the authors and not necessarily those of the Commonwealth of Australia. The Commonwealth of Australia accepts no responsibility for any loss, damage or injury resulting from reliance on any of the information or views contained in this publication

The Institute for International and Economic Policy (IIEP) also assisted with funding for a research assistant.

Abstract

Cash transfer programs are a popular social protection tool in developing countries that aim, among other things, to improve education outcomes in developing countries. The debate over whether these programs should include conditions has been at the forefront of recent policy discussions. This systematic review aims to complement the existing evidence on the effectiveness of these programs in improving schooling outcomes and help inform the debate surrounding the design of cash transfer programs. Using data from 75 reports that cover 35 different studies, the authors find that both conditional cash transfers (CCTs) and unconditional cash transfers (UCTs) improve the odds of being enrolled in and attending school compared to no cash transfer program. The effect sizes for enrollment and attendance are always larger for CCTs compared to UCTs but the difference is not statistically significant. When programs are categorized as having no schooling conditions, having some conditions with minimal monitoring and enforcement, and having explicit conditions that are monitored and enforced, a much clearer pattern emerges whereby programs that are explicitly conditional, monitor compliance and penalize non-compliance have substantively larger effects (60% improvement in odds of enrollment). Unlike enrollment and attendance, the effectiveness of cash transfer programs on improving test scores is small at best. More research is needed that examines longer term outcomes such as test scores and, more generally, evaluating the impacts of UCTs.

1. Introduction

Increasing educational attainment around the world is one of the key aims of the Millennium Development Goals (MDGs). The second MDG states that by 2015 “children everywhere, boys and girls alike, will be able to complete a full course of primary schooling,” while the third MDG aims to “eliminate gender disparity in primary and secondary education preferably by 2005 and at all levels of education no later than 2015.” Improved education is critical for decreasing poverty and inequality, as well as improving a host of other welfare measures (see Glewwe and Kremer 2006 for a review of the literature).

While schooling rates have been increasing globally - of 163 developing countries, 47 have achieved universal primary education - there are still many countries where they remain low (World Bank 2009). In two-thirds of sub-Saharan African countries, more than 30 per cent of primary students who start school are expected to drop out before they reach the last grade of primary education (UNESCO 2011). Secondary schools in many countries remain expensive, and even at the primary school level expenses on uniforms and other related necessities can make the cost of school prohibitive for poor households. Moreover, unlike universal access, universal completion cannot be achieved without ensuring household demand for education (Bruns, Mingat and Rakotomalala 2003).

Conditional cash transfers (CCTs) provide cash support, generally to poor households, conditional on certain behaviors on the part of the household. In the case of CCTs aimed at improving schooling, the condition for the eligible households to receive the regular cash payments might be to enroll school-aged children in school or for the children to keep their attendance rates above a certain threshold. Thus, CCT programs seek to encourage increased demand for schooling through an “income effect,” by increasing the income of the household, and a “substitution effect,” by decreasing the opportunity cost of schooling. In the majority of cases, these programs are run by the government, with a smaller set of programs (largely pilot programs or experiments) run by NGOs or other smaller organizations. The transfers are generally made to the mother in the household, with certain programs targeted specifically at other designated groups. The programs generally last until the household no longer has an eligible recipient, which typically happens for one of three possible reasons: the household may exceed the poverty threshold for eligibility, children may grow older than the upper age limit for the program, or benefits may be suspended because the conditions have not been met. Most CCT programs have used a means- or proxy-means-testing method to target beneficiaries. These methods consist of a narrow approach of assessing individual or household thresholds and criteria in order to determine who should receive benefits (Tesliuc, del Ninno, and Grosh 2009).

Unconditional cash transfers (UCTs) also target certain groups, although eligible households need not be poor. Furthermore, they are not tied to any particular behaviors on the recipient individuals’ or households’ parts, and thus provide cash payments to everyone in the eligible target population. Old age pension programs and child support grants are the most common forms of UCTs, but some programs also target orphans and vulnerable children. The main difference between these programs

and schooling CCTs is that UCTs do not explicitly specify any behavioral conditions to receive payments and thus act only through an income effect.ⁱ

The main argument for UCTs is that the key constraint for poor people is simply lack of money (for example, because of credit constraints), not knowledge, and thus they are best equipped to decide what to do with the cash (Hanlon, Barrientos and Hulme 2010). Additional income would allow them to make different investments in health and education, among other things. If the household's behavior is optimal (privately and socially) in the first place, then attaching a condition to the cash transfer will cause costly distractions to households who need the cash and will distort their behavior away from the optimal. Moreover a CCT, by attaching a condition, may exclude segments of the population who, for one reason or another, do not comply with program rules and may equally be in need of cash transfers.

On the other hand, there are three main arguments for attaching conditions to cash transfers. The first reason is the existence of a market failure that causes suboptimal levels of education among school-age children, even from a private point of view. Such failures can arise due to lack of information (for example, parents or teenagers do not know the true value of schooling), differences in discount rates (for example, parents discount future consumption at a higher rate than their children), or intra-household bargaining problems (for example, parents not valuing girls' education or community norms keeping families from sending girls to school). In such cases, CCTs can change household behavior towards a privately optimal investment in children's schooling. The second reason is that investments in education, even if privately optimal, may be below the socially optimal level because, for example, of positive externalities arising from the education process (for example, workers with colleagues with more education may be more productive or higher education levels may lead to lower crime rates, and so on). The final argument is one of political economy, whereby redistributive policies may be much more palatable to the taxpayers if transfers are seen to be 'rewarding' socially desirable behaviors rather than being simply 'handouts' (Fiszbein and Schady 2009).

A large and empirically well-identified body of evidence has demonstrated the ability of CCTs to raise schooling rates in the developing world (Schultz 2004; de Janvry et al. 2006; among many others). Due in large part to the high-quality evaluation of Mexico's PROGRESA program, CCT interventions have become common in Latin America and have spread to other parts of the world. Over 37 developing countries have implemented a form of CCT program (Fiszbein and Schady 2009; Grosh, et al. 2011), and in some cases multiple programs have functioned at the same time within the same country. CCTs are not limited to developing countries either – for example, Opportunity NYC was a three-year pilot CCT program in New York City, USA. The number of evaluations that assess the effects of UCTs on schooling is substantially smaller, but growing. Whether examining the old age pension program in South Africa, or the child support grants also in South Africa, studies find that UCTs improve schooling outcomes among children, along with other outcomes (see, for example, Duflo 2003).

The ultimate impact of a cash transfer program on schooling outcomes will depend on a number of moderating factors such as monitoring of the condition, transfer size, recipient of the transfer, baseline enrollment rate, and so on. It is important to note that although this systematic review aims to compare the relative effectiveness of UCTs and CCTs as two clearly distinct program types, in reality there exists a

range of programs that goes from pure UCTs to fully monitored and enforced CCTs. The secondary analysis adopted in this study is aimed to address the somewhat fuzzy nature of these programs by taking this heterogeneity into account while assessing the impact of each program.

The literature assessing the effectiveness of CCT programs on schooling is large. The list of CCT programs and references found in the review “Conditional Cash Transfers: Reducing Present and Future Poverty” (Fiszbein and Schady, 2009) contains data on over 40 programs and hundreds of references. It’s fair to say that CCTs are one of the most studied programs in development economics. In addition to the detailed narrative review of the evidence in Fiszbein and Schady (2009), there are a number of other reviews, including Parker et al (2008) and Adato and Bassett (2012). Rigorous evaluations of UCT interventions that schooling impacts form a smaller, but recently growing, set of studies. Hanlon, Barrientos and Hulme (2010) provide a review of these programs.

Despite the interest in the relative effectiveness of these two types of programs, the literature directly comparing them is limited. Until a few years ago, what we knew on the topic was limited to evidence from natural experiments due to glitches in program implementation (de Brauw and Hoddinott 2008; Schady and Araujo 2008) or structural models of household behavior (Bourguignon, Ferreira, and Leite 2003; Todd and Wolpin 2006). All of these studies concluded that the conditions played an important role in improving outcomes in schooling – over and above the income effect. Two studies in Latin America – Paxson and Schady (2010) and Macours, Schady, and Vakis (2008) – also show behavioral changes in the spending patterns of parents and households that they argue to be inconsistent with changes in household income alone.

The ideal experiment to identify the marginal contribution of the condition in cash transfer programs – that is, a randomized controlled trial with one treatment arm receiving *conditional* cash transfers, another one receiving *unconditional* transfers, and a control group receiving *no* transfers – has only recently been conducted. There are now four such studies in Burkina Faso, Malawi, Morocco and Zimbabwe, with varying success in implementation. These studies are all included in our meta-analysis and are also discussed narratively in the concluding section.

The objective of this review is to synthesize the evidence on the relative effectiveness of CCT and UCT programs in improving schooling outcomes in low- and middle-income countries (LMICs).ⁱⁱ It does so by identifying existing studies – experimental and quasi-experimental evaluations of both schooling CCT and UCT programs – to assess the overall effects on enrollment, attendance, and test scores for each type of intervention using a systematic review approach. A systematic review attempts to synthesize all the empirical evidence that meets pre-specified eligibility criteria to provide an answer on a given research question (Higgins and Green 2011). A systematic review may or may not include a meta-analysis, which is the practice of combining the results of individual studies include in the review into an overall statistic – an effect size. In addition to providing mean effect sizes with standard errors for each type of intervention, the review also presents effect sizes for four main subgroups: boys and girls; and primary and secondary schools. The study also analyzes the heterogeneity of effects across a number of important design parameters. These include the intensity of monitoring and enforcement of conditions in CCT programs (priming or labeling in UCT programs); transfer size; baseline enrollment rate; the

identity of the household member receiving the transfers; the frequency of transfers and whether the program is in piloting phase or fully scaled-up. Sensitivity analysis includes analysis of the results according to whether the underlying studies are experimental or not; the risk of bias for each study as assessed by the authors; and whether or not the studies are published in peer-reviewed journals. The authors also assess the presence of publication bias and conduct robustness checks to ensure that the conclusions drawn from the systematic review are not biased.

Previously, no such systematic review had been conducted. There is, however, one comprehensive review on conditional cash transfers and a similarly comprehensive review on unconditional cash transfers (Fiszbein and Schady 2009; Hanlon, Barrientos and Hulme 2010). Neither of these reviews attempts a systematic comparison of the two types of transfers. There are also two reviews that focus on the impact of CCTs on health outcomes (Gaarder, Glassman and Todd, 2010 and Lagarde, Haines and Palmer 2007), one that looks at the impact of CCTs on education outcomes (Saavedra and García 2012), and a recent review on the impact of cash transfers and employment guarantee schemes on poverty (Hagen-Zanker et al. 2011). This study adds to the existing literature by comparing the relative effectiveness of CCT and UCT programs on schooling outcomes. Given the relative paucity of studies making such direct comparisons of CCT and UCT interventions, the authors hope that a systematic review making indirect comparisons will be an important contribution that is useful to policymakers.

The remainder of this paper is structured as follows. Section 2 briefly discusses the methods. Section 3 presents the results. Section 4 provides a discussion and Section 5 concludes.

2. Methods

This analysis uses a systematic review approach that is outlined in detail in the Methods Appendix. This section provides a short summary of some of the key points discussed in the Methods Appendix.

2.1. Criteria for considering studies for this review

Studies included in this systematic review include both experimental (randomized control trials) and quasi-experimental designs with a controlled comparison. The search included studies in English, Portuguese and Spanish. The search was also restricted to publications after 1997, which corresponds with the onset of PROGRESA/Oportunidades. Limiting the search to this start-date allows for a more comparable group of CCT and UCT interventions. This analysis is restricted to low and middle-income countries (as defined by the World Bank), where the majority of schooling CCT and UCT programs are implemented, with no other explicit population exclusion criteria.

The population of focus in this study is those targeted by either UCT or schooling CCT programs. Schooling CCT programs typically, although not exclusively, target poor families with school-aged children, while UCT programs generally target a broader spectrum of the poor population. Thus, the entire set of eligible interventions is largely targeted at disadvantaged populations. Outcome variables

are restricted to children of ages 5-22 to cover impacts related to primary and secondary school education.. Early childhood development and higher education outcomes are beyond the scope of this review.

This review focuses on schooling outcomes often cited in the cash transfer literature. Our immediate outcomes of interest were school attendance and school enrollment. Our final outcome of interest was test scores

2.2. Search methods for identification of studies

Databases searched are included in Table 1. We restricted all searches to papers published since 1997. The initial search was completed on April 18, 2012. The search terms used are also listed in Table 1. In addition to the database search, we also contacted researchers who have published on the topic of conditional or unconditional cash transfers and asked for references on unpublished work to minimize publication bias in our summary. We also asked researchers to indicate if they or other colleagues are working on relevant studies, in order to allow us to incorporate ongoing work not yet published. Our advisory panel also sent additional references. We also reviewed websites of organizations working in the field to search for relevant grey literature and contacted relevant researchers in these organizations. In addition we conducted hand searches of the past five years (January 2008-April 2012) of a selection of key journals. We then investigated the bibliographies uncovered through the first two steps to check for other citations that might meet the search criteria.

Finally, given the delay of approximately one year between the end of the initial search and the submission of the final draft, we updated our references with all new eligible references the study team was aware of as of 30 April 2013.

2.3. Data Analysis

We analyze three outcomes: enrollment, attendance and test scores. For each outcome we constructed an effect size and its standard error which takes into account the effect of clustering.

In this systematic review, there are multiple layers of information, which requires us to make the following definitions before data synthesis can be exposed clearly. We define an **intervention** to be a UCT or a CCT. We define a **study** to be a different version of a UCT or a CCT (or in a few experiments a UCT **and** a CCT) implemented in different places. For many of these studies, there are multiple publications (journal articles, working papers, technical reports, and so on). We refer to these as **reports**. In our meta-analysis, the unit of observation is the **study**. In our review we construct one effect size per study for the overall effect on any of our three outcome variables and for each subgroup (gender and level of education). This implies that all the different estimates within a study have to be combined into one effect size per subgroup.ⁱⁱⁱ We do this, for each subgroup, by synthesizing and summarizing multiple effect sizes within each report, then again synthesizing and summarizing those combined effect sizes from different reports within a study. We analyze the data in Stata.

Our analysis looks at whether the effect size is significantly different between the CCT arm and the control, the UCT arm and the control, as well as between the CCT arm and the UCT arm. We look at overall results, as well as conduct the analysis by subgroup. We also undertook multivariate meta-regression analysis to explore whether the results are moderated by the following variables: transfer amount, mean follow-up enrollment rate in the control group, transfer recipient, and number of transfers per year, and whether the study is a pilot program or a national scaled-up intervention. To test the robustness of our conclusions regarding the methodological quality of the studies, we undertook sensitivity analysis, where we excluded all studies with high risk of bias. In addition, we calculated pooled ES by restricting the sample to randomized studies.

3. Results

3.1. Results of the search

The initial database search retrieved a total of 4,167 publications, of which 4,041 were deemed ineligible by reviewing the citation and abstract. The reasons these reports were deemed ineligible are as follows (and presented in Table 2A): duplicate (n=1489), not an experimental or quasi-experimental study (146), did not fit language or date requirements (230), of no relevance to the study question (2176). The majority of ineligible reports either were completely unrelated to the study question or the duplicate of another report. This left us with a total of 126 publications from the initial database search that would move to the full article review stage. In addition, a number of additional eligible references were found through website searches (four references), hand searches (eight references), and searches of other relevant systematic reviews (17 references) following the procedure indicated above. This resulted in a total of 155 reports that moved to the full article review stage.

Out of these 155 reports, eight could not be downloaded leaving us with a total of 147 full articles downloaded. Out of these 147 reports, 75 were initially deemed ineligible largely due to not being a primary study (21) and not having an impact estimate (16) (see Table 2B for a full list of reasons). This left us with 72 eligible references.^{iv}

There were then two final sets of checks. Initial comments from reviewers on the included studies identified five additional studies for inclusion, with an additional six added at the final draft stage. An additional 12 reports were deemed older versions of an eligible paper. This left us with a final total of 75 included reports and 95 excluded reports.

3.2. Description of the reports and studies

3.2.1. Included Reports and Studies

We included 75 reports that are listed in section 14.1. These 75 reports consist of 33 journal articles, 27 working papers, 10 technical reports, four dissertations, and one unpublished manuscript (Table 3, Panel

A). Additional details of each publication are listed in Online Appendix Table B. Thirty-five of the reports utilize experimental methods, while the remaining 40 utilize quasi-experimental methods. 61 of the reports focus on CCTs, 10 on UCTs, and four on direct UCT/CCT comparisons. Publication year ranges from 2000-2013, with the number of reports growing in recent years. Figure 2 illustrates the increasing trend in publications.

In terms of outcome measures, 67 reports include enrollment, 17 include attendance, and 12 include test scores. Online Appendix Table C indicates what outcomes are reported for each report.

These 75 publications correspond to 35 studies in 25 countries.^v On average there are 2.17 reports per study. Characteristics of these 35 studies are summarized in Table 3, Panels B and C, with details of each study listed in Online Appendix Tables D1 and D2. Overall, the included studies include 26 CCT programs, five UCT programs, and four programs that contain both CCT and UCT components. The 35 programs consist of 19 programs in Latin America and the Caribbean, eight programs in Asia, and eight programs in Africa. Out of these 35 programs, nine were pilot programs and 12 had random assignment. Thirty-two of the studies target both genders, with three focusing only on girls.

The average follow-up enrollment rate in the control group is 79 per cent. The average transfer size is calculated to be 5.7 per cent of annual household income, with households receiving on average 8.2 transfers per year. The annual per person cost of the program is approximately \$351.

Given that the focus of this review is on the relative effectiveness of CCTs versus UCTs, it is important to understand the underlying characteristics of each of the three types of studies in our review.

CCT Programs (26)

CCT programs are by far the most common intervention type included in our review. Approximately 30 per cent of these studies included random assignment. CCTs are also frequently national programs with 22 out of 26 at the national level. The CCT programs are all either in Latin America and the Caribbean (18) or Asia (8).

UCT Programs (5)

There are five UCT programs, four of which are in Africa and one in Latin America and the Caribbean. None of these programs use random assignment and four of them are national programs.

CCT/UCT Experiments (4)

Direct comparisons of CCTs and UCTs are relatively new to the literature with one journal article in 2011, another one in 2013, and a working paper and an unpublished manuscript in 2013. These are an important sub-group of studies for our analysis as they are experiments comparing CCTs to UCTs within a country. Each study includes at least one CCT and one UCT arm. All four of these studies are pilots in sub-Saharan Africa and they are all randomized.

As discussed in Section 3.1.3, a binary categorization of cash transfer programs as CCT and UCT proves somewhat inadequate (Özler 2013). There are differences between programs in terms of what the rules are on paper (the *de jure* conditions), the information disseminated to the beneficiary population, the ‘priming’ of the value of schooling for children, monitoring of the conditions, and enforcement of any penalties or sanctions. While this multi-dimensional space is hard to navigate and define linearly, it is nonetheless possible to categorize interventions along a continuum from purely unconditional to explicitly conditional (with rules monitored and enforced). We present such an attempt here and use it to complement our meta-analysis of schooling effects of these programs by intervention type. We categorize the studies in this review according to the intensity of the conditionality with respect to social marketing campaigns, monitoring, and enforcement as follows (the four randomized CCT/UCT programs are counted twice):^{vi}

0. UCT programs unrelated to children or education – *such as Old Age Pension Programs* (2)
1. UCT programs targeted at children with an explicit aim of improving education – *such as Kenya’s CT-OVC or South Africa’s Child Support Grant* (2)
2. UCTs that are conducted within a rubric of education – *such as Malawi’s SIHR UCT arm or Burkina Faso’s Nahouri Cash Transfers Pilot Project UCT arm* (3)
3. Explicit conditions on paper or education encouragement, but not monitored or enforced – *such as Ecuador’s BDH or Malawi’s SCTS* (8)
4. Explicit conditions, (imperfectly) monitored, with minimal enforcement – *such as Brazil’s Bolsa Familia or Mexico’s PROGRESA* (8)
5. Explicit conditions with monitoring and enforcement of **enrollment** condition – *such as Honduras’ PRAF-II or Cambodia’s CESSP Scholarship Program* (6)
6. Explicit conditions with monitoring and enforcement of **attendance** condition – *such as Malawi’s SIHR CCT arm or China’s Pilot CCT program* (10)

3.3. Excluded Reports

We excluded 95 reports, which are listed in section 15. The three main reasons for exclusion were not a primary study (21), no impact estimate (16) and earlier version of an eligible paper (12). An additional eight references could not be downloaded. For a full list of the reasons why studies were excluded refer to Table 2B.^{vii}

3.4. Risk of bias in included studies

Table 5 presents the summary results from the risk of bias assessment. Online Appendix Table E presents the paper level risk of bias results.

1. *Selection bias and confounding*: Overall 18 out of the 75 reports (24 per cent) completely address this issue. While there are many RCTs with excellent identification strategies, many of them fail to discuss attrition rates (or have very high attrition) or have issues with sample size. In addition, many of the quasi-experimental designs do not provide the level of detail necessary to completely determine the quality of the study. Moreover, for some categories of quasi-experimental design the best possible ranking is ‘unclear,’ as indicated in the risk of bias tool.

2. *Spillovers, cross-overs and contamination:* Thirty-five (47 per cent) of reports adequately address this issue. While many of the included reports use clustering of some sort, the majority of reports are from national programs where contamination is possible. Moreover, quite a few studies report contamination of the treatment group, where individuals assigned to control were actually treated.
3. *Outcome reporting:* All but one paper adequately addresses the issue of outcome reporting, and there is no evidence of selective reporting. The review currently does not consider papers that report more objective measures of education outcomes (for example spot checks at schools) as superior to self-reports of enrollment. As reports increasingly collect more objective measures of schooling, this would also be worth taking into consideration.
4. *Analysis reporting:* The majority of reports take an appropriate approach when conducting the analysis, with 55 (73 per cent) reports addressing this sufficiently. The main reason a report was deemed of lower quality for this category was the failure to report the necessary tests for quasi-experimental methods.
5. *Other risks of bias:* Other risks of bias show up when either the author has to create data they do not have access to or baseline data is collected retrospectively. Out of the 75 reports, 64 (85.3 per cent) are assessed to have no other risks of bias.

Utilizing the above categories, we categorize the reports as low risk of bias, medium risk of bias and high risk of bias. Forty-eight per cent of the reports are categorized as low risk of bias, 24 per cent as medium risk of bias and 28 per cent as high risk of bias. Ultimately the separation into the three categories is largely driven by whether the underlying intervention used random assignment, as well as whether reports that used quasi-experimental designs discuss all relevant features of the approach. Spillovers and contamination are also frequently a problem due to the fact that the majority of the interventions are national level programs.

It is also interesting to look at the risk of bias by the three intervention types. Reports on CCTs, which form the majority of the reports, have a similar distribution to the overall risk of bias (52 per cent low, 23 per cent medium and 25 per cent high). Reports on UCTs are on average of lower quality (10 per cent low, 30 per cent medium, and 60 per cent high), a result that is largely driven by the fact that the UCT programs are frequently not designed in a manner that makes it easy to undertake impact analysis. Finally, the four studies that utilize direct CCT vs. UCT interventions are the highest quality group (75 per cent low, 25 per cent medium and zero per cent high) largely because these are all pilot RCTs, allowing the researcher to have more control over the study design.

3.5.Synthesis

3.5.1. Quantitative synthesis

Enrollment

Figure 3 and Table 5 present the main findings of the systematic review for 32 studies and 35 effect sizes with an overall effect size for school enrollment. The ES is reported in changes in the odds of being enrolled in school. Twenty-seven of these studies are defined to be CCTs while the remaining eight are UCTs. For the 35 CCT and UCT interventions studies combined, the pooled ES is 1.36 (95% CI 1.24-1.48), meaning that the odds of children being enrolled in school is 36 per cent higher among children in households offered cash transfers compared with children in households who were not offered to participate in a cash transfer intervention. The effect is statistically significant at the 99% level (p -value <0.001).

Two panels in this figure present the same analysis by binary intervention type. The top panel indicates that UCT interventions significantly increase the odds of being enrolled in school (compared to a control group receiving no intervention), but the size of the effect is lower than the pooled ES described above (OR 1.23, 1.08-1.41). Of the eight studies categorized as UCT interventions, the ES ranges between 1.04 and 1.59, only one of which (Morocco's Tayssir Pilot) is significantly higher than one. I-squared of 52.2 per cent (p -value=0.041) suggests that only half of the variation in ES is due to heterogeneity between studies.

The bottom panel of the same figure, presenting the findings for CCTs, indicates that the odds of being in school are also higher in these studies (1.41, CI 1.27-1.56). In contrast to the UCT studies, most of the variation in ES is explained by between-study heterogeneity (I-squared 86.5, p -value <0.001). The tests of heterogeneity are consistent with the fact that most of the ES for UCT studies are near zero (with the exception of Morocco's Tayssir), while there is a wider distribution of effect sizes among CCT studies. The ORs range from a statistically insignificant 0.72 (Turkey's SRMP), to a very significant 4.36 (Nicaragua's RPS).

We tested whether the pooled ES for UCTs is different than that for CCTs by running a meta-regression with a CCT dummy as the explanatory variable. The coefficient estimate for the CCT dummy, presented in Table 5, is 1.15 (95% CI 0.94-1.41), indicating that while the odds of being enrolled in school under CCTs is 15 per cent higher than under UCTs, this difference is not statistically significant (p -value=0.183).

Given the small number of UCT studies in our analysis, one or two studies can influence the overall findings. We have mentioned earlier that Morocco's Tayssir Pilot UCT arm ended up *de facto* imposing an enrollment condition at the outset of the program, while Ecuador's BDH is categorized as a CCT even though it neither monitored nor enforced any conditions. If these studies are re-categorized as a CCT and UCT respectively, the OR between CCT and UCT interventions becomes 1.21 (p -value=0.080). The results are very similar if these 'fuzzy' interventions are excluded from the analysis (OR 1.22, p -value=0.083).

The fact that we are trying to force complex interventions with varying design parameters into a binary straightjacket introduces noise into the data and may reduce the precision of our estimates regarding the absolute and relative effectiveness of these interventions. Rather than defining each study as a CCT or a UCT and then checking the robustness of the findings with respect to those definitions (as we have done in the previous paragraph), an alternative approach is to analyze the data with respect to a

categorical variable that describes the intensity of the conditionalities in each study and conduct the meta-analysis by that variable. Such a variable that takes on discrete values from zero to six was described in Section 4.2.1, with zero indicating an unconditional cash transfer program, and a six indicating an explicitly conditional CT program that is monitored and enforced, with all other programs falling on a linear continuum in between.

The results of the analysis using this ‘intensity of conditionality’ variable are presented in Table 5 and Figures 4 and 5. In Figure 4, we can see that the effect sizes are close to zero in studies with no (or low intensity) conditionalities, but increase steadily as the intensity of the conditionalities rise. The numbers in Table 5 suggest that the OR is 1.05 (p-value=0.63) when the intensity variable is equal to zero and increases by 6.7 per cent (p-value=0.011) for each unit increase in the intensity of conditionalities. An easier way to visualize these effects is presented in Figure 5, which presents a forest plot with three broader categories: (i) no schooling conditions (intensity=0, 1, or 2); (ii) some schooling conditions with no enforcement or monitoring (intensity=3 or 4); and (iii) explicit schooling conditions monitored and enforced (intensity=5 or 6). The pooled ES (95% CI) of these groups are 1.18 (1.05-1.33), 1.25 (1.10-1.42), and 1.60 (1.37-1.88), respectively. The 95% CI for studies with no conditions and studies with conditions monitored and enforced do not overlap. Meta-regression analysis (not shown here) indicates that the group with explicit conditions monitored and enforced has a significantly higher pooled ES than either of the other two groups (as well as those two other groups combined).

Figure 5 also indicates that the variation due to between-study heterogeneity is zero among studies with no schooling conditions, while those figures are much higher at 87 per cent and 80 per cent, respectively for studies with some conditions and studies with explicit conditions. The effect sizes lie between 1.04 and 1.31, 0.72-1.96, and 1.05-4.36 for these three groups of studies, respectively.

Before we move to subgroup analysis on enrollment, we touch on one final point with regards to moderators of these effect sizes by categories of interventions. The meta-regression analysis presented in Table 5 indicates that the intensity of conditionalities variable explains only a small fraction of the between-study heterogeneity in effect sizes: the residual variation due to heterogeneity after accounting for this variable is 82.6%. This means that other design parameters, such as transfer size, identity of the transfer recipient, the frequency of transfers, or mean enrollment rate in the control group, may explain some of the variation in effect sizes across these studies. Table 5 also presents the findings from such a multivariate meta-regression analysis. Surprisingly, none of the added design elements – transfer size (as a percentage of baseline household income), whether the transfer is given to the mother (or another woman in the household), whether transfers are monthly, bimonthly, quarterly, or annual, whether the study is a pilot program, or the level of school enrollment in the control group – has a significant effect in moderating the pooled ES. Furthermore, the addition of these moderators does little to change the moderating effect of the ‘intensity of conditionalities:’ the coefficient estimate for that variable is 1.08 (p-value=0.005). The residual variation due to between-study heterogeneity is 74 per cent, meaning that unobserved variation in other aspects in the design of these programs account for the considerable variation in effect sizes.

Figures 6-9 present forest plots among four pre-defined subgroups: boys, girls, primary schools, and secondary schools (results summarized in Table 6). Not all studies present findings by these subgroups as evidenced by the substantially smaller number of studies for each analysis. This selection means that there may be a correlation between relative effect sizes and the likelihood of presenting subgroup estimates due to unobserved heterogeneity. Hence, the findings here should be treated with caution and interpreted as suggestive. Figures 6 and 7 show the subgroup analysis among boys and girls, respectively, by the binary intervention type. The pooled ES for UCT and CCT among boys are 1.28 (0.97-1.69) and 1.55 (1.28-1.86), respectively. The same figures are 1.32 (1.10-1.60) and 1.64 (1.43-1.88) among girls. These findings are consistent with the overall effects on enrollment presented above, although the larger pooled ES are again a reminder of the fact that these are a select group of studies and that the findings should be treated with caution. Figures 8 and 9 present pooled ES for CCT interventions at the primary and secondary levels, respectively. These plots are limited to CCTs because there is only one UCT study that reports findings by schooling level. While the number of studies in this analysis is small (six studies for primary and 10 studies for secondary schools), the pattern that emerges suggests that CCT programs are only effective in increasing enrollment at the secondary level. The pooled ES at the primary and secondary levels are 1.04 (0.98-1.11) and 1.31 (1.16-1.48), respectively.

Attendance

A smaller number of studies assess intervention effects on school attendance rather than enrollment. There are 16 studies (one UCT, eleven CCTs, four UCTs/CCTs) that report attendance in our meta-analysis giving us a total of 20 effect size estimates as reported in Figure 10 and Panel A of Table 7. As with enrollment, both types of interventions significantly increase the likelihood that children are attending school. The ORs in UCT and CCT groups are 1.42 (1.18-1.70) and 1.65 (1.37-1.99), respectively. Again, like enrollment, meta-regression analysis suggests that the likelihood of attending school increases with the intensity of the conditionalities (OR 1.082, p-value=0.132).

Subgroup analysis for boys (limited to two CCT-UCT experiments) and girls (limited to three CCT-UCT experiments and Cambodia's JFPR) suggests that the effects of CCTs and UCTs are similar for boys but that CCTs may be more effective for girls than UCTs for increasing attendance (analysis not shown here). Two experiments (Malawi's SIHR and Burkina's Nahouri) find significant differences in attendance impacts between CCT and UCT arms among girls, while one experiment (Morocco's Tayssir) finds no such differences.

Test Scores

Our meta-analysis is limited to five studies (three of which are CCT-UCT experiments) that report standardized test scores, which is reported in Figure 11 and Panel B of Table 7. Neither type of intervention has a significant effect on test scores. The pooled effect sizes are 0.04 and 0.08 standard deviations, respectively, for UCT and CCT interventions – small effects by any interpretation. When we examine the three experiments directly comparing the effects of CCT versus UCT treatment arms, we find that while Malawi's SIHR and Burkina's Nahouri programs both find small but statistically significant differences between CCT and UCT arms with respect to test scores (favoring CCTs), Morocco's Tayssir

study finds the opposite. The findings are similar in studies that are in this systematic review but excluded from the meta-analysis here: no consistent effects on test scores for CCTs or UCTs. It's fair to conclude that the effects of these interventions on student achievement are small at best. More studies are needed to discern any differences between intervention types.

3.6. Analysis of heterogeneity

Tables 5-7 also show the I-squared and τ^2 measures of heterogeneity. For UCT studies the I-squared of 52.2 per cent suggests that only half of the variation in ES is due to heterogeneity between studies. In contrast to the UCT studies, most of the variation in ES for CCTs is explained by between-study heterogeneity (I-squared 86.5). The tests of heterogeneity are consistent with the fact that most of the ES for UCT studies are near zero (with the exception of Morocco's Tayssir), while there is a wider distribution of effect sizes among CCT studies.

Table 5 also indicates that the variation due to between-study heterogeneity is zero among studies with no schooling conditions, while those figures are much higher at 87 per cent and 80 per cent, respectively for studies with some conditions and studies with explicit conditions.

3.7. Sensitivity analysis

Table 8 Panel A and B replicate the overall enrollment results restricting first to RCTs and second to low and medium risk of bias studies. The results for both the UCT and the CCT are relatively stable regardless of what sub-set of reports you look at. For the UCT group the effect size is 1.23 for the overall sample, 1.31 restricting to RCTs, and 1.26 restricting to low and medium risk of bias studies. For CCTs, the effect sizes are 1.41, 1.43 and 1.43.

3.8. Publication bias

A large share (56 per cent) of the reports in this review are from the grey literature, meaning that they are working papers, technical reports, or unpublished manuscripts. Furthermore, the PIs have contacted numerous authors and experts to unearth reports that might have otherwise been missed. The search was recently updated with new reports. All these factors should minimize publication bias.

However, studies with non-results may still have not been written up, which may cause publication bias. Table 8, Panel C presents the pooled ES for enrollment when we restrict our sample to published articles only. All effect sizes are still significantly different than the control group. The difference between the CCTs and UCTs is both larger in this group and statistically significant at the 90% level (1.25, p-value=0.065), which may be due to publication bias or a difference in the quality of evidence (risk of bias) that was not accounted for in our analysis. It may also reflect the fact that this is a rapidly growing literature (see Figure 2) and many of the papers are still in the working paper stage but are likely to eventually become journal articles.

To further examine the possibility of publication bias, Figure 12 presents a funnel plot of effect sizes against the standard errors – with the sample restricted to CCT studies only.^{viii} The figure does suggest that high variance studies with smaller effect sizes may be missing from the sample of studies included in this review. While the usual interpretation for this finding is 'small study bias,' none of the studies in our review really fit this criterion, as they are evaluations of mostly large national programs. It is

possible that studies with other sources of error (such as measurement error) causing high variance with small effect sizes are missing.

The important question is whether such potential bias is likely to alter our main findings. To assess this, we conduct a cumulative, random-effects meta-analysis and present our findings in Figure 13. The figure, which is sorted by the weight of each study in the RE analysis, indicates that the cumulative ORs increase as the weight of the studies decline. More importantly, however, the ORs stabilize around 1.35 after about 18 studies (out of 27). This implies that even if we excluded studies with high variance (those at the bottom right corner of the funnel plot in Figure 12) from our analysis, the OR (and its 95% CI) would not be substantively altered.

4. Discussion

4.1. Summary of findings

Tables 9 and 10 provide a summary of the key findings of this review on the impact of CCTs and UCTs on enrollment, attendance and test scores. We have data on enrollment from 32 studies, of which three have direct comparisons of CCTs and UCTs, leaving us with 35 effect sizes. Our findings suggest that both CCTs (OR 1.41, 95% CI 1.27-1.56) and UCTs (OR 1.23, 95% CI 1.08-1.41) have a significant effect on enrollment. These results indicate that CCTs increase the odds of a child being enrolled in school by 41 per cent and UCTs increase the odds by 23 per cent. We do not find a significant difference when comparing CCTs to UCTs (OR 1.15, 95% CI 0.94-1.42].

The binary categorization of these programs into CCT versus UCT ignores the fact that there is a great deal of variation in the intensity of the conditionality. In order to exploit this heterogeneity we construct a variable that takes on discrete values from zero to six with zero indicating an unconditional cash transfer program and six indicating an explicitly conditional CT program that is monitored and enforced, with all other programs falling on a linear continuum in between. The results of this analysis show the OR is 1.05 (p-value=0.63) when the intensity variable is equal to zero and increases by 6.7 per cent (p-value=0.011) for each unit increase in the intensity of the condition.

If we instead group the conditionality variable into three broader categories (i) no schooling conditions, (ii) some schooling conditions with no enforcement or monitoring and (iii) explicit schooling conditions monitored and enforced we find odds ratios as follows: 1.18 (95% CI 1.05-1.33), 1.25 (95% CI 1.10-1.42), and 1.60 (95% 1.37-1.88), respectively. The 95 per cent CI for studies with no conditions and studies with conditions monitored and enforced do not overlap. Meta-regression indicates that outside of the intensity of the condition none of the other measured design elements – transfer size (as a percentage of baseline household income), whether the transfer is given to the mother (or another woman in the household), whether transfers are monthly, bimonthly, quarterly, or annual, whether the study is a pilot program, or the level of school enrollment in the control group – has a significant effect in moderating

the overall effect size. Even controlling for these variables the residual variation due to between-study heterogeneity is high at 74 per cent, meaning that unobserved variation in other aspects in the design of these programs account for the considerable variation in effect sizes.

A smaller number of studies assess intervention effects on school attendance rather than enrollment. There are 16 studies (one UCT, eleven CCTs, four UCTs/CCTs) that report attendance in our meta-analysis giving us a total of 20 effect size estimates as reported in Figure 10 and Panel A of Table 7. As with enrollment, both types of interventions significantly increase the likelihood that children are attending school. The ORs in UCT and CCT groups are 1.42 (1.18-1.70) and 1.65 (1.37-1.99), respectively. Again, like enrollment, meta-regression analysis suggests that the likelihood of attending school increases with the intensity of the conditionalities (OR 1.082, p-value=0.132).

Our meta-analysis is limited to five studies (three of which are CCT-UCT experiments) that report standardized test scores. The pooled effect sizes are small at 0.04 (95% CI: 0.041-0.121) and 0.08 (95% CI: -0.002-0.162) standard deviations, respectively, for UCT and CCT interventions – small effects by any interpretation. It's fair to conclude that the effects of these interventions on student achievement are small at best. More studies are needed to discern any differences between intervention types.

4.2.Overall completeness and applicability of evidence

Our review included 35 studies, of which four included direct comparisons of CCTs and UCTs. There are a lot more evaluations of CCTs (26) than UCTs (5) with 4 direct comparisons of CCTs versus UCTs (all in Africa). The CCT programs are all either in Latin America or Asia, while the UCTs are almost exclusively in Africa. The fact that these different types of interventions are taking place on different continents limits the applicability of evidence. The small number of UCTs in the sample, and the fact that none of them utilized randomization, is a limitation of this review.

The majority of the reports analyze enrollment, with fewer investigating attendance and fewer still analyzing test scores. Thus, we are limited in our ability to assess the impact of cash transfers on longer term outcomes. There are also a number of reports that include test score data that cannot be included in the meta-analysis either because they do not report standardized effects or because they conduct test scores on a select sample (that is, only those in school), thus biasing their impact estimates.

Design elements of CCTs and UCTs are likely important for the overall effect of the program, although our analysis suggests that only the intensity of the conditions matters. While we are able to code a number of other important design features, it would have been useful to code many more, but many of the reports do not provide sufficient detail on these features.

Finally, it would have been useful to conduct a more rigorous assessment of cost, other than simply looking at transfer size. Very few reports discuss cost at all, let alone provide the detail needed to conduct cost effectiveness analysis.

4.3.Quality of the evidence

The quality of reports varied, with 21 of the 75 ranked as high risk of bias, 36 as low risk of bias, and 18 of medium risk. The large number of high quality reports is encouraging and reflects the strong evaluation culture around cash transfer programs, particularly CCTs. In fact, when we focus on CCTs only, only 25 per cent of the reports are high risk of bias, compared with 60 per cent of UCTs. None of the reports that utilize a direct comparison of UCTs and CCTs are considered high risk of bias.

We are able to estimate effect sizes for 35 studies for enrollment and 20 studies for attendance. For test scores, however, we can only estimate effect sizes for eight studies, thus limiting our ability to discuss final outcomes

4.4.Potential biases in the review process and limitations of the review

While the review team has minimal experience with meta-analysis, the analysis was done with frequent guidance from the Methods Coordinating Group of The Campbell Collaboration. There are a number of limitations of this review. First, many of the references used in the analysis utilize quasi-experimental designs that sometime rely on imperfect identification thus leading to more uncertainty regarding the measurement of the treatment effect. Second, economic papers tend not to report the exact information needed for effect size calculations, so certain assumptions have to be made in order to calculate the effect size. Third, outside of the four studies that experimented with CCT and UCT arms, the UCT and CCT programs tend to be operating in different countries and contexts, which considerably complicates comparisons. Finally, and perhaps one of the main limitations of this review is that there are far fewer studies evaluating UCTs than CCTs.

4.5.Agreements and disagreements with other studies or reviews

This is the first systematic review that we know of that compares the relative effectiveness of CCTs and UCTs in improving education outcomes. The one systematic review that does review the impact of CCTs on education outcomes (Saavedra and Garcia 2012) also finds an overall positive and significant effect of CCTs on education outcomes. Note that although this previous review was published recently (February 2012), our analysis includes an additional 23 references for CCTs suggesting this is a fast growing body of evidence. It is difficult to make a direct comparison with this review as they use a different approach to calculate their effect sizes.

5. Conclusions

5.1.Implications for policy and practice

In this review, we examine the effects of cash transfer interventions on schooling outcomes. We do this by synthesizing effect sizes across studies of cash transfer interventions by both defining interventions as a binary variable (CCT or UCT) and by a more delineated discrete variable (intensity of conditionalities). As discussed earlier, these programs are complex interventions with many design parameters at the disposal of policymakers, including conditions, transfer size, identity of the transfer recipient, transfer frequency, monitoring, and enforcement. While the title of this systematic review refers to conditional and unconditional cash transfer programs, many programs simply can't be neatly defined as one or the other.

Our main finding is that these programs improve the odds of being enrolled in and attending school. Using our binary categorization of cash transfer interventions, we find that both CCTs and UCTs are effective in improving school participation. The effect sizes for enrollment and attendance are always larger for CCT programs but the statistical significance of this difference varies slightly by the choice of categorization of studies.

However, when programs are categorized as having no schooling conditions, having some conditions with minimal monitoring and enforcement, and having explicit conditions that are monitored and enforced, a much clearer pattern emerges. While interventions with no conditions or some conditions that are not monitored have some effect on enrollment rates (18-25 per cent improvement in odds of being enrolled in school), programs that are explicitly conditional, monitor compliance and penalize non-compliance have substantively larger effects (60 per cent improvement in odds of enrollment).

The findings of relative effectiveness on enrollment in this systematic review are also consistent with experiments that contrast CCT and UCT treatments directly. Two experiments in Malawi and Burkina Faso both find larger effects on enrollment for CCTs than UCTs. The third such experiment from Morocco finds no significant differences between the CCT and the UCT arms, but, as discussed earlier, the difference between these treatment arms was dulled by the fact that UCTs were conditional on enrollment at the outset of the program.^{ix}

Unlike enrollment and attendance, the effectiveness of cash transfer programs on test scores is small at best. While the latest experiments suggest a modest improvement in test scores, likely due to better measurement, these effects are still quite small (<0.1 SD). It seems likely that without complementing interventions, cash transfers are unlikely to improve learning substantively.

Our study has some limitations. First and foremost, a comprehensive assessment of CCT and UCT programs would ideally involve a comparison of welfare effects rather than the limited question of schooling effects addressed in this review. Beyond the direct costs of conditions to eligible households (and to the implementing agencies), CCT programs can introduce significant trade-offs compared with a counterfactual of cash transfers with no strings attached. For example, Baird, McIntosh, and Özler (2011) shows that while CCTs outperformed UCTs in terms of improving schooling outcomes in Malawi, UCTs were more effective in preventing teen pregnancies and marriages – due to the fact that girls who dropped out of school and lost their CCT payments were more likely to get married and pregnant than girls who dropped out of school in the UCT arm. Policymakers would be well-advised to keep such trade-

offs in mind while designing cash transfer programs. However, this does not mean that assessing the relative effectiveness of CCTs and UCTs with respect to schooling outcomes is meaningless. Given the additional cost of imposing conditions in cash transfer programs, establishing whether CCTs are indeed more effective than UCTs in improving schooling outcomes is a necessary, but not sufficient, condition in justifying their *raison d'être*. Such evidence can also help policymakers assess the kinds of trade-offs described above.

Second, there simply are too few rigorous evaluations of UCTs. For example, currently there is limited evidence that UCTs that have nothing to do with children or schooling have any effect on schooling outcomes, but the fact that there are only two such studies in our review (Old Age Pensions in South Africa and Social Security Reform in Brazil) suggests that more rigorous assessments of such programs are needed before any conclusions can be drawn with confidence.

Third, while our review was able to precisely identify effect sizes by intervention type and identify one influential moderator (intensity of conditionalities), most of the heterogeneity in effect sizes remains unexplained. This suggests that unobserved design elements, local implementation modalities, as well as context and culture may cause considerable variation in expected effect sizes.^x While our effect sizes are precise, when we predict the range of outcomes for a future trial (taking into account the heterogeneity in effect sizes across studies), the estimated predicted intervals include an odds ratio of one (that is, no effect) for both types of interventions.

Finally, we originally hoped that this review would also provide some evidence on the relative cost-effectiveness of CCTs versus UCT programs. Unfortunately, there was very limited cost data available, making it difficult to assess the cost effectiveness in a rigorous manner.

5.2. Implications for research

There are a number of limitations to this review that we feel could be mitigated with additional research. First, in order for systematic reviews and meta-analyses to be a useful tool in economics, authors need to do a more thorough job of reporting details of the study design, as well as the numbers necessary for the effect size calculation. In particular, authors need to report the follow-up mean of the outcome variable in the control group for binary outcomes, as well as either pooled standard deviations or standard deviations and sample sizes for each group. In addition, authors should always report an overall impact alongside sub-group estimates to help with comparisons across studies. Finally, authors should include a basic breakdown of costs.

Second, many studies still rely on self-reported outcome measures, although this seems to be changing. The risk of bias due to self-reporting is something that needs to be assessed further, but existing evidence suggests that it may be large, especially in experiments – due to differential reporting between treatment arms (Baird and Özler 2012; Barrera-Osorio et al. 2011). Future studies would be better served to objectively measure enrollment, attendance, and learning outcomes.

Third, further research is needed on evaluating UCTs to increase the evidence base for this set of studies and allow for more confidence in comparing the relative effectiveness of CCTs and UCTs. Moreover, additional research is needed on programs that contain both a CCT and a UCT component. Along this same line, more replication studies of CCT and UCT programs in different settings are needed to understand the role of country context in influencing results and to build a broader body of evidence.

Acknowledgements

We would like to thank the International Development Coordinating Group of the Campbell Collaboration for their assistance in development of the protocol and draft report. We would especially like to thank John Eyers and Emily Tanner-Smith as well as anonymous referees for detailed comments that greatly improved the protocol. We would also like to thank David Wilson for help with the effect size calculations. We would also like to thank Josefina Durazo, Reem Ghoneim (with financial support from the Intsitute for International Economic Policy at George Washington University), and Pierre Pratley for research assistance.

Methods Appendix

Criteria for considering studies for this review

Types of Studies

Eligible studies included experimental (randomized control trials) and quasi-experimental designs with a controlled comparison.^{xi} Quasi-experimental designs required a cross sectional and/or longitudinal comparison (for example, controlled before and after, cross-sectional, interrupted time series, parallel cohort, and regression discontinuity design). For quasi-experimental designs we indicate the method of analysis used to control for endogeneity of program placement (for example, regression discontinuity designs, instrumental variables, matching, and difference in difference). These causal identification strategies are investigated as a potential source of effect size variation and discussed in the risk of bias assessment.

The search included studies in English, Portuguese and Spanish. The search was also restricted to publications after 1997, which corresponds with the onset of PROGRESA/Oportunidades. Limiting the search to this start-date allows for a more comparable group of conditional and unconditional interventions.

We do not exclude studies based on publication status. Comments, op-eds, summaries or media briefings, purely qualitative studies and non-experimental observational studies are excluded.

Types of Participants

This analysis is restricted to low and middle-income countries (as defined by the World Bank), where the majority of schooling CCT and UCT programs are implemented, with no other explicit population exclusion criteria. The population of focus in this study is those targeted by either UCT or schooling CCT programs. Schooling CCT programs typically, although not exclusively, target poor families with school-aged children, while UCT programs generally target a broader spectrum of the poor population. Thus, the entire set of eligible interventions is largely targeted at disadvantaged populations. Outcome variables are restricted to children of ages 5-22 to cover impacts related to primary and secondary school education, including vocational training. Early childhood development and higher education outcomes are beyond the scope of this review.

Types of interventions

While the title of this systematic review suggests a binary distinction between CCT and UCT interventions, in practice, the distinctions are not as clear-cut. As a first pass, we categorize an intervention to be a CCT if it contains one or more conditions explicitly related to schooling – at least on paper. UCT interventions are defined as those with no explicit conditions related to schooling (for example the Old Age Pensions in South Africa), but excluding contributory pensions and disability grants. The included UCT interventions contain child support grants, non-contributory pensions, and old age pensions, as well as cash transfer programs that are explicitly unconditional. However, not all interventions neatly fall into these two categories. For example, Bono Desarrollo de Humano (BDH) in Ecuador was intended to be a CCT program (with rules on paper, some community meetings suggesting the same, TV and radio campaigns encouraging beneficiaries to invest in their children’s education and health) but ended up being unconditional in the sense that no schooling-related conditions were monitored or enforced. Malawi’s Social Cash Transfer Scheme (SCTS) is called a UCT but provided a ‘schooling attendance’ bonus to families with school-age children. Yet another intervention (the UCT arm of Morocco’s Tayssir pilot program) ended up being *de facto* conditional on school enrollment (because the program was overseen by the Ministry of Education and run through the headmaster of the local schools, who enrolled children in school while they were being enrolled in the program). In such “fuzzy” cases, we have opted to stick with the original designation of the program (that is, CCT for BDH and UCT for SCTS and Tayssir), but as we discuss later, we also construct another variable that delineates the existence of conditions, any social marketing surrounding the importance of investing in children’s schooling, and the monitoring and enforcement of conditions – ranging from a pure UCT (such as Old Age Pension Programs) to a pure CCT (such as Malawi’s Schooling, Income and Health Risk study, where conditions were explicit, monitored closely, and enforced swiftly). This variable takes on discrete values between zero and six, with zero being assigned to a pure UCT and a six being assigned to a pure CCT intervention.^{xii}

Eligible comparison groups

Eligible comparison groups include both a direct comparison between a CCT intervention and a UCT intervention, as well as comparisons between a CCT and a control and a UCT and a control. The control group must either be constructed using an experimental design or using one of the quasi-experimental methods listed in 3.1.1

Types of outcome measures

This review focuses on schooling outcomes often cited in the cash transfer literature. Our immediate outcomes of interest were school attendance and school enrollment. Our final outcome of interest was test scores.^{xiii} For enrollment and attendance, we include studies that utilize self-reported data from household surveys, as well as more objective data such as administrative data, data from surveys of the school, and unannounced school visits. Data on test scores used in our meta-analysis come from studies using tests that were designed to evaluate the impact of that particular program on learning and were administered at the participants' homes. A few other studies utilize school-based tests, such as end-of-year exams, which are likely to suffer from selection bias. Those studies are included in the systematic review but only discussed narratively in the result sections and excluded from the meta-analysis of effect sizes on test scores.

Search methods for identification of studies

Electronic Searches

Databases searched are included in Table 1. We restricted all searches to papers published since 1997. The initial search was completed on April 18, 2012. The search terms used are also listed in Table 1.

Other Searches

In addition to the database search, we also contacted researchers who have published on the topic of conditional or unconditional cash transfers and asked for references on unpublished work to minimize publication bias in our summary. We also asked researchers to indicate if they or other colleagues are working on relevant studies, in order to allow us to incorporate ongoing work not yet published. Our advisory panel also sent additional references.^{xiv}

We also reviewed websites of organizations working in the field to search for relevant grey literature. These organizations included: African Development Bank, Asian Development Bank, Australian Agency for International Development (AusAID), Department for International Development, Inter-American Development Bank, International Food Policy Research Institute, International Institute for Impact Evaluation, Pan American Health Organization, Swedish development agency, UNDP, USAID, UNICEF, UNESCO, World Bank, and the WHO. Along with searching the websites, we contacted researchers in these organizations involved in cash transfer programs for further documentation.

In addition we conducted hand searches of the past five years (January 2008-April 2012) of the following journals: American Economic Journal: Applied Economics, American Economic Review, Economic

Development and Cultural Change, Journal of Development Economics, Journal of Development Effectiveness, Quarterly Journal of Economics, World Development, and World Bank Economic Review.

We then investigated the bibliographies uncovered through the first two steps to check for other citations that might meet the search criteria.

Finally, given the delay of approximately one year between the end of the initial search and the submission of the final draft, we updated our references with all new eligible references the study team was aware of as of 30 April 2013. This included six new publications, and four working papers that had become journal articles.

Data collection and analysis

Selection of studies

The selection of studies was based on the search methods outlined above. The search took the following steps:

1. *Step 1:* Two research assistants searched the above listed databases for the above listed search terms contained in reference titles, abstracts or keywords for publications 1997 and later. The research assistants also undertook bibliographic back-referencing, hand searches in relevant journals, website searches, and discussions with researchers. The PIs resolved any discrepancies arising from this process. Studies meeting our inclusion criteria were downloaded into the bibliographic management software RefWorks. At this point, duplicate records of the same report were deleted.
2. *Step 2:* Two research assistants independently read the abstract to make sure the studies met the geographic criteria of being implemented in a low or middle-income country as defined by the World Bank.
3. *Step 3:* Two PIs independently read the abstract, introduction, methodological sections and tables and retained references that met the inclusion criteria as set out above.

Any inconsistencies between the two researchers were then discussed and resolved by looking at the details of the manuscripts.

Data extraction and management

Online Appendix Table A provides a list of the data extracted from the papers. Once the papers were saved in RefWorks, data was extracted into a Microsoft Excel file by two people and subsequently entered into Stata. Any disagreements were debated and a final decision agreed upon. Subsequent data analysis was conducted in Stata.

Assessment of risk of bias in included studies

We utilized the risk of bias tool developed by the International Development Coordinating Group (IDCG) secretariat to assess risk of bias. This tool has been developed to assess the risk of bias for a range of quasi-experimental studies, as well as experimental studies. The tool is attached as Online Appendix F. For each of the five categories listed below we coded the paper as 'Yes' if it addresses the issue, 'No' if it did not, and 'Unclear' if it was unclear. We then aggregated to an overall risk of bias as Low, Medium or High based on an aggregation across the five categories as follows:

1. Low Risk of Bias: 'Yes' for four or five categories
2. Medium Risk of Bias: 'Yes' for three categories
3. High Risk of Bias: 'Yes' for two or less categories

The five categories used to assess risk of bias are briefly discussed below, and then presented in detail in Online Appendix F. The categories are as follows:

1. *Selection bias and confounding*: addresses the issue of the design of the program. This category addresses the issue of whether or not the allocation was free from any sources of bias or whether sources of bias were adequately corrected for with an appropriate method of analysis. For details of the coding see Online Appendix F.
2. *Spillovers/cross-overs/contamination*: addresses the issue of spillovers from the treatment to the control group. This variable is coded as 'Yes' if spillovers are unlikely from the treatment to the control group through geographic or social separation. The variable is coded as 'No' if spillovers are likely through, for example, individual level randomization and are not addressed appropriately in the manuscript. The variable is coded as 'Unclear' if spillovers and contamination are not addressed.
3. *Outcome reporting*: addresses the issue of whether analysis of all relevant outcomes was reported or whether there appears to be selection in reporting. Coded as 'Yes' if all relevant outcomes reported, 'No' if selective reporting is apparent, and 'Unclear' if not specified in the paper.
4. *Analysis reporting*: this category is coded as 'Yes' if the authors utilize a credible analysis method to deal with attribution given the data available, and is coded as 'No' otherwise. The category is coded as 'Unclear' if not enough detail is given to ascertain whether they are utilizing the most appropriate method.
5. *Other risks of bias*: this category is coded as 'No' if there are other risks of bias present in the report. These may include data on the baseline collected retrospectively, information collected using an inappropriate instrument or a different instrument/at different time/after different follow up period in the control and in the treatment group, and so on. This is the most subjective of the five categories.

We utilized the risk of bias tool for sensitivity analysis and to understand the overall quality of the data.

Measures of treatment effect

Measures of treatment effects come from three different types of studies: CCT versus control, UCT versus control, and, for four experimental studies, CCT versus UCT. For these latter set of studies, a separate effect size for CCT and UCT (each compared with the control group of no intervention) is constructed. We analyze three outcome measures, described below.

Enrollment

For a binary outcome variable, such as enrollment, the standard practice is to calculate odds ratios (OR) using follow-up means of success (p , or enrollment rate) and failure rates ($1-p$, or share not enrolled) and its standard error using sample sizes. Thus, under ideal circumstances, that is, for a randomized controlled trial (RCT) conducted at the individual (and not cluster) level that reports unadjusted means at baseline and follow-up (or just at follow-up if baseline balance is not an issue), enrollment rates and the sample sizes in the treatment (T) and control (C) groups are sufficient to be able to calculate effect sizes (ES) in the form of ORs and their standard errors (SE).

However, the studies covered under this systematic review do not fit neatly into this ideal picture. First, and most importantly, the treatment is assigned at the community level rather than the individual level. The studies reviewed here are virtually all cluster RCTs (or use other causal identification methods to assess interventions implemented at the cluster level) and use survey sampling that also employ clustering. This implies that even when follow-up enrollment rates and sample sizes are available, which is often not the case for the reports eligible for this systematic review, the standard errors cannot be calculated using the usual formula. This is because the standard errors of the ES have to take into account the intra-cluster correlation in the outcome variable; so calculating SE without clustering would produce smaller standard errors and overstate the precision of the estimates.

Second, while a good number of studies are cluster RCTs, many of the studies reviewed here use other plausible causal identification strategies. These include propensity score matching (PSM), difference-in-difference estimation (DD), PSM DD, triple difference estimation (DDD), regression discontinuity (RD), and, very rarely, cross-sectional estimates with community fixed effects and a rich set of control variables. Hence, both the RCTs and other studies deemed eligible for inclusion under this systematic review use regression models to estimate the effect of cash transfers on educational outcomes. These regressions most often take the form of linear probability models, but some studies also use probit and logit models. They almost always utilize adjustments for baseline covariates (even in the case of RCTs to protect against chance imbalance and to improve precision). The fact that many of the studies are not RCTs and the fact that almost all studies adjust impact estimates for the inclusion of baseline covariates (and fixed effects as necessary) implies that the simple differences in follow-up enrollment rates between T and C are not the same as the impact estimates reported by these regression models.

Hence, in this review, to calculate a pooled ES, the follow-up mean enrollment rate in C and the impact estimate on enrollment of T obtained from the regression model are required. These provide us with a raw success rate in C and a covariate-adjusted success rate in T.^{xv} Using these two figures the OR can

easily be calculated. However, there is still the issue of calculating the standard error of the OR, which, as mentioned earlier, cannot be calculated using the usual formula.

To tackle this issue, we have decided to follow Wilson (2011) and convert the logged OR, or $\ln(\text{OR})$, into a standardized effect size, d . As the logistic distribution is similar to the normal distribution and the logged ORs conform to the logistic distribution, we can convert each $\ln(\text{OR})$ into a d using the following formula:

$$d = \ln(\text{OR})/1.814.^{\text{xvi}}$$

Then, the standard error of d can be calculated using the standard error of the coefficient estimate for the treatment indicator from the appropriate regression model as follows:

$$SE_d = d/z,$$

where z is either a z - or t -test associated with the treatment effect from the regression model.

Hence, our main methodology to calculate ES and its standard error in each study is to code the follow-up enrollment rate in C; calculate the (covariate adjusted) follow-up enrollment rate in T by adding the impact estimate from the regression model to the enrollment rate in C; calculating $\ln(\text{OR})$ using these two figures; converting the logged OR into a d using the linear adjustment described above; and calculating the standard error of d using the t -stat (or z -stat) associated with the impact estimate from the regression model.¹

Note that if the regression model is a linear probability model (LPM) or a probit model reporting marginal effects, this is interpreted as a percentage point change in the treatment group over and above the control mean at follow-up. If the regression model is a logit, then the reported estimate is the logged OR.

Attendance

Attendance is measured in two different ways in the studies considered in this review. First, researchers may be using data that were collected during random visits to the schools/classrooms, which were used to discern whether study participants were attending school that day or not. The outcome is a binary variable that takes the value of 'one' if the student is present that day and 'zero' otherwise. The average of this variable in a treatment or control group is then the 'success' rate, or the percentage of students that were present on the randomly chosen day of school visit.

Other studies ask the student (or his/her parents) how many days he/she missed school for a given period, such as past week, past two weeks, past two months, and so on. In this case, the outcome is a discrete variable that takes on values between zero and the maximum number of school days during the

¹ For reports that analyzed program effects on the likelihood of dropping out of school instead of being enrolled in school, we similarly calculated the implied enrollment levels in the treatment and control groups at follow-up.

recall period. The percentage of days the students have missed are then also averaged into a 'success rate,' which is the mean share of days the students were at school during the recall period.

Given that both types of data are ultimately converted into a 'success' rate that is bounded between zero and one, we treat attendance as if it was a binary outcome and calculate the standardized effect size d (and its standard error) in exactly the same way as we do for enrollment, that is, by calculating odds ratios, converting them into standardized effect sizes, and using the t-statistics associated with the treatment effect from the regression analysis to calculate the standard error and variance for d . Constructed this way, the attendance variable can be interpreted as the probability of a randomly chosen student being present at school on a randomly chosen day during the evaluation period.

Test Scores

The third and final outcome we consider in our systematic review of the relative effectiveness of CCT and UCT programs is achievement test scores. Ten studies report impact estimates on achievement tests, such as mathematics, reading, writing, vocabulary, and cognitive skills. The test scores are continuous variables, reported on different scales.

Five studies report impacts using standardized test scores obtained from tests that were developed specifically to evaluate the impact of those particular interventions on learning and administered to children at their homes. We restrict the meta-analysis of test scores to these studies. When there is more than one test in a report, such as mathematics and french, they are combined to create one 'synthetic effect size' per report. Details of effect sizes that are synthesized and summarized within and across reports within a study are discussed in detail further below.

Of the five other studies, all CCT evaluations, four do not provide sufficient information to calculate a proper standardized effect size.^{xvii} Another study uses end-of-year tests administered at school, which has a high risk of bias. We exclude these studies from the meta-analysis, but discuss their findings narratively in the results section.

Unit of analysis issues

The CCT and UCT interventions are always implemented at a cluster level, while the unit of analysis is always the individual. Therefore, as mentioned above, when calculating standard errors of the effect size we take account of the clustering.

Dealing with missing data

We contacted study authors for information on missing data and updates. However, we were still left with missing data.

Within the sub-sample of reports included in our review, there was a good deal of heterogeneity in terms of what was reported. As a result, we had to make a number of assumptions in order to calculate effect sizes:

1. In many instances, the mean enrollment rate at follow-up is not reported for the sample or for a sub-group. In such cases, the follow-up rate is assumed to be equal to the baseline mean (that is, no change over time in C).^{xviii} If information on the time trend is available (for example, from the text or figures in the study or from another study for the same country), then this information is used as appropriate. If no baseline or follow-up enrollment rate is available for C (that is, the study only reports an impact estimate without any reference to a control mean at baseline or follow-up), then the study is excluded (*ineligible - ES cannot be calculated*).
2. Similarly, in many instances, baseline (or follow-up) means are available for the entire sample, but not for the sub-groups for which impact estimates are provided. In such cases, we assign the sub-groups the same mean as the overall sample, that is, assume that the success rates are equal for, say, different grades, boys and girls, or urban and rural areas.
3. Sometimes, the studies report enrollment rates for the entire sample, that is, T & C, instead of reporting them separately for T and C. In such cases, we assume that the sample mean at baseline is equal to the mean in C.
4. Some studies report standard errors, others t-stats, and others p-values. In all of these cases, it is possible (to a close approximation) to calculate the t-stat needed to calculate the standard error of the standardized ES. In some rare cases, studies only report stars to indicate statistical significance at the 1 per cent, 5 per cent, and 10 per cent levels. In such cases, the t-stat is calculated using the most conservative estimate of the p-value (that is, for 0.01, 0.05, and 0.10, respectively).
5. In rare cases, the follow-up enrollment rate in C plus the (covariate-adjusted) impact estimate from the regression model is larger than one. In such cases, it is not possible to calculate an OR. They are replaced by 0.999.
6. One study (Khandker, Pitt, and Fuwa 2003) uses variation in exposure to the CCT program across geographic areas rather than a treatment indicator. In that case the one-year ATE is used as a substitute for the program's impact.
7. To calculate the share of days the students were in attendance during a given recall period, we assumed that the schools were in session five days during the past week, and 22 days during the past month unless these figures were provided by the authors.

Assessment of heterogeneity

We report estimates of the between-studies variance component τ^2 , the Chi-squared test of heterogeneity, and the I-squared statistic. We attempt to analyse the factors explaining heterogeneity through moderator analysis using meta-regression models that include intervention design parameters as independent variables.

Assessment of reporting biases

Reporting bias is assessed by re-estimating the pooled effect sizes using only journal articles, as well as through the use of funnel plots and cumulative meta-analysis. These issues are also discussed in Section 5 under the strengths and limitations of this review.

Data synthesis

In this systematic review, there are multiple layers of information, which requires us to make the following definitions before data synthesis can be exposed clearly.

We define an **intervention** to be a UCT or a CCT. While there are a few countries with multiple interventions in our review, the large majority of countries have one intervention in our sample. There are many design elements that make up an intervention, such as transfer size, the identity of the transfer recipient, the dissemination, monitoring, and enforcement of the conditions imposed on the beneficiaries, and so on. Each country implements its particular cash transfer intervention in different ways.

Hence, we define a **study** to be a different version of a UCT or a CCT (or in a few experiments a UCT **and** a CCT) implemented in different places. For example, Mexico's PROGRESA is a study. So are Mexico's Oportunidades, Malawi's Social Cash Transfer Scheme, South Africa's Old Age Pension Scheme, and so on. For example, in our meta-analysis of enrollment, there are 35 different studies in 24 countries.

For many of these studies, there are multiple publications (journal articles, working papers, technical reports, and so on). We refer to these as **reports**. For example, there are 15 reports that assess the impact of Mexico's PROGRESA CCT on schooling outcomes, while three reports assess the effect of Brazil's Bolsa Escola. The number of reports per study depends on the availability of data (particularly whether data were collected specifically to evaluate the impact of that study), as well as whether the study used random assignment to determine treatment and control groups. For example, all reports on PROGRESA utilize the same data source while each report on Bolsa Escola uses a different data set.

To add to the complexity of these layers, each report may contain multiple estimates for the same outcome. For example, there may be enrollment effects from multiple follow-up surveys (assessing shorter- and longer-term effects); from multiple achievement tests (such as English, Spanish, and Math); using multiple estimation techniques (with and without baseline controls, nearest neighbor matching vs. one-to-one matching), and so on. Furthermore, some studies report effects only for subgroups (such as by age or urban/rural or grade or sex) but report no overall effect.

In our meta-analysis, the unit of observation is the **study**. This means that we would like to construct one effect size per study for the overall effect on any of our three outcome variables and for each subgroup (if reported). This implies that all the different estimates within a study have to be combined into one effect size per subgroup.^{xix} We do this, for each subgroup, by synthesizing and summarizing (explained below) multiple effect sizes within each report, then again synthesizing and summarizing those combined effect sizes from different reports within a study.

We create *synthetic effects* when the effect sizes are not *independent* of each other. This is the case when there are multiple effects reported for the same sample of participants – such as effects of

enrollment in 2001 and 2002; effects on Math and English tests, effects using two different estimation techniques. In such cases, the effects are combined using a simple average of each effect size (ES) and the variance is calculated as the variance of that mean with the correlation coefficient r (between effect sizes being combined) assumed to be equal to 1.^{xx}

However, when two or more ES are independent of each other, we create *summary effects*. These are mostly cases where a report provides ES by subgroups but does not provide an overall estimate. To combine these estimates into an overall estimate (or an estimate for a pre-defined subgroup), we utilize a random effects (RE) model.^{xxi}

Once effect sizes have been combined as described above to produce one overall ES per outcome and one for each subgroup (as available), multiple reports within studies are summarized and synthesized in the same manner to produce one ES per outcome per study. For example, as all reports for PROGRESA use the same data set, the reports are combined to produce a synthetic effect for that study. On the other hand, reports are summarized using a RE model for Bolsa Escola, where each report uses a different sample.

With ES being combined in this manner up to the study level, the effect sizes per study can be considered independent, meaning that they can be analyzed using standard tools of meta-analysis.

We use a RE model (using the 'metan' command in Stata) to produce forest plots.^{xxii} Because we have ES and its standard error already calculated per study, these two variables are fed into 'metan' the pooled ES by intervention type.^{xxiii} We report the overall ES, as well as ES by intervention type, their confidence intervals, as well as tests of heterogeneity. To test whether the pooled ES for CCT interventions is significantly different than that for UCT interventions, we employ meta-regression analysis ('metareg' in Stata). *Metareg* with no independent variables yields identical results to metan when method of moments is used to estimate the between-study variance and standard normal distribution is used to calculate p-values and confidence intervals. Similarly, we employ metareg to conduct moderator analysis with multiple moderators used as explanatory variables. We employ the *metafunnel* and *metacum* commands in Stata to analyze the possibility of publication (small study) bias in our systematic review.

Subgroup analysis

We undertook sub-group analysis according to the following characteristics of the recipients: gender and level of education (primary versus secondary). We also undertook multivariate meta-regression analysis to explore whether the results are moderated by the following variables: transfer amount, mean follow-up enrollment rate in the control group, transfer recipient, and number of transfers per year, and whether the study is a pilot program or a national scaled-up intervention.

Sensitivity analysis

To test the robustness of our conclusions regarding the methodological quality of the studies, we undertook sensitivity analysis, where we excluded all studies with high risk of bias. In addition, we also calculated pooled ES by restricting the sample to randomized studies.

Tables

Table 1: Databases and Search Terms

Databases	Search Terms
<p>ABI Inform Complete, ADOLEC, African-Wide, African Journals OnLine (AJOL), British Education Index, CAB Direct, Center for Reviews and Dissemination, The Cochrane Library, EBSCO, Econlit, Eldis, Effective Practice and Organization of Care Group (EPOC) Reviews, ERIC, FRANCIS, German Education Index, Google Scholar, Healthcare Management Information Consortium, International Bibliography of the Social Sciences (IBSS), IDEAS, Inter-Science Latin American and Caribbean Health Sciences Literature (LILACS), JOLIS library catalogue - International Monetary Fund, World Bank and International Finance Corporation, MEDCARIB, NBER, Pan American Health Organization (PAHO) Library Catalogue, PAIS International, POPLINE, ProQuest Dissertation and Theses Database, PsycInfo, Scielo, ScienceDirect, Scopusxxiv, Social Science Research Network (SSRN), Sociological Abstracts, Web of Science, WHO's Global Health Library, WHOLIS (World Health, Organization Library Database)</p>	<p>ONE OF:</p> <p>Cash adj3 transfer* OR non-contributory adj pension* OR noncontributory adj pension* OR non adj contributory adj pension* OR child adj support adj3 grant* or child* adj grant* OR old adj age adj pension*</p> <p>AND ONE OF (A) or (B)</p> <p>Immediate Outcomes (Educ*) OR (School* AND (attend* OR enrol* OR dropout* OR participat* OR complet*))</p> <p>Longer Term Outcomes Cognitive, test score, grade attended OR level adj3 attain*?, grade point average, grade adj3 progress*, grade promotion, grade adj3 (repetition or repeat*), return* to education, standardized test or standardised test</p>

Table 2A: Reference screening procedure to obtain full pdf sample

Phase 1: Database Search	Number
Total references downloaded	4167
Total ineligible references	4041
Reason ineligible:	
Duplicates	1489
Not experimental or quasi experimental	146
Did not fit language or date requirements	230
Dropped relevance	2176
Phase 1: Total eligible references	126
Phase 2: Additional eligible sources from other search methods	
Website search	4
Hand search	8
Other systematic reviews	17
Phase 2: Total eligible references	29
Total eligible references for full review	155

Table 2B: Reference screening procedure to obtain analysis sample

Phase 3: Full review of articles	Number
Total full articles to be reviewed	155
Unable to access	8
Total full articles downloaded	147
Total ineligible references	75
Reason ineligible	
Developed country	1
No relevant education outcome	9
No impact estimate	16
Not a cash transfer program	8
Not a primary study	21
Research design does not meet requirements	11
Duplicate	5
Not enough information to calculate effect size	4
Phase 2: Total eligible references	72
Phase 3: Final checks	
Advisory board and other expert reviewers	5
Old version of an eligible paper	8
Phase 3: Total eligible references	69
Phase 4: New references since end of original search	
New papers found since original search	6
Working papers updated with journal article (working paper version moved to excluded)	4
Total Eligible References	75

Table 3: Characteristics of analysis sample

Panel A: Reference level characteristics: (N=75)		
	Number	%
Publication type:		
Journal article	33	44.00%
Working paper	27	36.00%
Technical Reports	10	13.33%
Dissertation	4	5.33%
Unpublished	1	1.33%
Reports effects on:		
Enrollment/Dropout	67	89.33%
Attendance	17	22.67%
Test Score	12	16.00%
Panel B: Study level characteristics, binary (N=35)		
	Number	%
UCT	5	14.29%
CCT	26	74.29%
UCT/CCT	4	11.43%
Regional Distribution		
Latin America and the Caribbean	19	54.29%
Asia	8	22.86%
Africa	8	22.86%
Female recipient	16	45.71%
Pilot Program	9	25.71%
Random Assignment	12	34.29%
Panel C: Study level characteristics, continuous (N=35)		
	Mean	Std
Control Follow-up Enrollment Rate	0.785	0.146
# of Reports per Study	2.17	2.360
Transfers per Year	8.24	4.020
Transfer amount (% of HH Income)	5.66	7.890
Annual per Person Cost (USD)	351	414

Table 4: Summary of Risk of Bias in Included Studies

Panel A: Summary of Risk of Bias by Category					
	Selection Bias and Confounding	Spillovers, cross-over, contamination	Outcome reporting	Analysis Reporting	Other Risks
Yes	18	35	74	55	64
Unclear	36	6	0	6	0
No	21	34	1	14	11

Panel B: Overall Assessment of Risk of Bias, by Intervention Type			
	Low Risk	Medium Risk	High Risk
Overall	48% (N=36)	24% (n=18)	28% (N=21)
CCT	52% (N=32)	23% (N=14)	25% (N=15)
UCT	10% (N=1)	30% (N=3)	60% (N=6)
CCT/UCT	75% (N=3)	25% (N=1)	0% (N=0)

Table 5: Summary of Effect Size for Enrollment, N=35

	Odds Ratio	P> z	95% Confidence Interval	τ^2	I-squared	Chi-squared Test
<u>Overall, CCT, UCT</u>						
Overall (vs. Control)	1.36	0.000	[1.24-1.48]	0.04	84.50%	219.19
UCT (vs. Control)	1.23	0.002	[1.08-1.41]	0.02	52.20%	14.64
CCT (vs. Control)	1.41	0.000	[1.27-1.56]	0.05	86.50%	193.15
CCT vs. UCT (regression)	1.15	0.183	[0.94-1.41]	0.04	84.12%	
<u>Condition Enforcement</u>						
No Schooling Condition	1.18	0.005	[1.05-1.33]	0.00	0.00%	1.14
Some Schooling Condition	1.25	0.001	[1.10-1.42]	0.04	87.20%	101.25
Explicit Conditions	1.60	0.000	[1.37-1.88]	0.06	80.60%	72.25
Intensity of Conditionality (regression)	1.07	0.011	[1.01-1.12]	0.04	82.60%	
<u>Meta-regression</u>						
Intensity of Conditionality	1.08	0.005	[1.02-1.14]	0.04	73.51%	
Transfer amount (% of HH Income)	1.00	0.502	[0.98-1.01]			

Control Follow-up Enrollment Rate	0.78	0.443	[0.41-1.48]	
Female Transfer Recipient	0.94	0.626	[0.74-1.20]	
Pilot Program	0.88	0.329	[0.67-1.14]	
Transfers per Year	1.01	0.348	[0.99-1.03]	

Table 6: Summary of Effect Size for Enrollment by Subgroup

	Odds Ratio	P> z	95% Confidence Interval	τ^2	I-squared	Chi-squared Test
<u>Panel A: Gender</u>						
<u>Boys (N=15)</u>						
Overall (vs. Control)	1.47	0.000	[1.26-1.72]	0.07	81.60%	76.29
UCT (vs. Control)	1.28	0.082	[0.97-1.69]	0.05	69.10%	9.70
CCT (vs. Control)	1.55	0.000	[1.28-1.86]	0.07	83.30%	59.99
CCT vs. UCT	1.21	0.292	[0.85-1.73]	0.07	81.35%	
Intensity of Conditionality (regression)	1.09	0.087	[0.99-1.19]	0.07	82.67%	
<u>Girls (N=19)</u>						
Overall (vs. Control)	1.55	0.000	[1.38-1.73]	0.04	71.30%	62.77
UCT (vs. Control)	1.32	0.004	[1.10-1.60]	0.02	47.20%	7.58
CCT (vs. Control)	1.64	0.000	[1.43-1.88]	0.04	73.80%	49.65
CCT vs. UCT	1.26	0.082	[0.97-1.65]	0.04	70.29%	
Intensity of Conditionality (regression)	1.06	0.117	[0.99-1.13]	0.0431	72.9%	
<u>Panel B: Schooling Level (CCT Only)</u>						

Primary (N=6)

CCT (vs. Control)	1.04	0.201	[0.98-1.11]	0.00	88.20%	42.26
Intensity of Conditionality (regression)	1.58	0.000	[1.31-1.90]	0.00	78.1%	

Secondary (N=10)

CCT (vs. Control)	1.31	0.000	[1.16-1.48]	0.02	89.20%	83.64
Intensity of Conditionality (regression)	1.09	0.122	[0.98-1.21]	0.02	88.62%	

Table 7: Summary of Effect Size for Attendance and Test Scores

	Odds Ratio	P> z	95% Confidence Interval	τ^2	I-squared	Chi-squared Test
Panel A: Attendance (N=20)						
<u>Overall, CCT, UCT</u>						
Overall (vs. Control)	1.59	0.000	[1.35-1.87]	0.09	91.80%	230.90
UCT (vs. Control)	1.42	0.000	[1.18-1.70]	0.00	0.00%	1.90
CCT (vs. Control)	1.65	0.000	[1.37-2.00]	0.10	93.60%	217.31
CCT vs. UCT (regression)	1.17	0.439	[0.79-1.74]	0.10	91.79%	
Intensity of Conditionality (regression)	1.08	0.132	[0.98-1.20]	0.09	90.97%	
Panel B: Test Scores (N=9)						
<u>Overall, CCT, UCT</u>						
Overall (vs. Control)	0.061	0.026	[0.007-0.115]	0.00	35.20%	10.79
UCT (vs. Control)	0.040	0.331	[-0.041-0.121]	0.00	21.30%	2.54
CCT (vs. Control)	0.080	0.056	[-0.002-0.162]	0.00	50.90%	8.15
CCT vs. UCT (regression)	0.046	0.470	[-0.080-0.173]	0.00	43.90%	

Intensity of Conditionality (regression)	0.019	0.293	[-0.016-0.055]	0.00	41.91%	
--	-------	-------	----------------	------	--------	--

Table 8: Sensitivity Analysis/Publication Bias of Effect Size for Enrollment

	Odds Ratio	P> z	95% Confidence Interval	τ^2	I-squared	Chi-squared Test
<u>Panel A: RCT (N=15)</u>						
<u>Overall, CCT, UCT</u>						
Overall (vs. Control)	1.40	0.000	[1.21-1.61]	0.06	90.00%	140.28
UCT (vs. Control)	1.31	0.015	[1.05-1.63]	0.03	68.70%	9.57
CCT (vs. Control)	1.43	0.000	[1.21-1.69]	0.05	90.80%	108.57
CCT vs. UCT (regression)	1.10	0.555	[0.81-1.49]	0.05	89.00%	
<u>Panel B: Low/Medium Risk of Bias (N=27)</u>						
<u>Overall, CCT, UCT</u>						
Overall (vs. Control)	1.38	0.000	[1.25-1.52]	0.04	87.20%	203.02
UCT (vs. Control)	1.26	0.004	[1.07-1.47]	0.02	61.90%	13.11
CCT (vs. Control)	1.43	0.000	[1.28-1.59]	0.04	88.70%	176.38
CCT vs. UCT (regression)	1.13	0.260	[0.91-1.42]	0.04	86.10%	
<u>Panel C: Journal Articles (N=17)</u>						

Overall, CCT, UCT

Overall (vs. Control)	1.34	0.000	[1.20-1.49]	0.03	70.40%	54.05
UCT (vs. Control)	1.15	0.014	[1.03-1.28]	0.00	0.00%	1.36
CCT (vs. Control)	1.47	0.000	[1.25-1.71]	0.04	77.90%	49.74
CCT vs. UCT (regression)	1.25	0.065	[0.99-1.59]	0.03	70.65%	

Table 9: Summary of Findings (Enrollment)

	Odds of Child Being Enrolled in School:	Statistically Significant?*	# Effect Sizes*	Comments
<u>CCT vs. UCT</u>				
Overall (vs. Control)	36% higher	Yes	35	Our analysis of enrollment includes 35 effect sizes from 32 studies. Both CCTs and UCTs significantly increase the odds of a child being enrolled in school, with no significant difference between the two groups. This binary distinction masks considerable heterogeneity in the intensity of the monitoring and enforcement of the condition. When we further categorize the studies, we find a significant increase in the odds of a child being enrolled in school as the intensity of the condition increases. In addition, studies with explicit conditions have significantly larger effects than studies with some or no conditions.
UCT (vs. Control)	23% higher	Yes	8	
CCT (vs. Control)	41% higher	Yes	27	
CCT (vs. UCT)	15% higher	No	35	
<u>Condition Enforcement</u>				
No Schooling Condition (vs. Control)	18% higher	Yes	6	
Some Schooling Condition (vs. Control)	25% higher	Yes	14	
Explicit Conditions (vs. Control)	60% higher	Yes	15	
Intensity of Condition	Increases by 7% for each unit increase in intensity of condition.	Yes	35	
Notes: We consider a study to be statistically significant if it is significant at the 95% level or higher. We use the term effect size here instead of study since the studies that directly compare CCTs and UCTs have two effect sizes in the analysis. All other studies have one.				

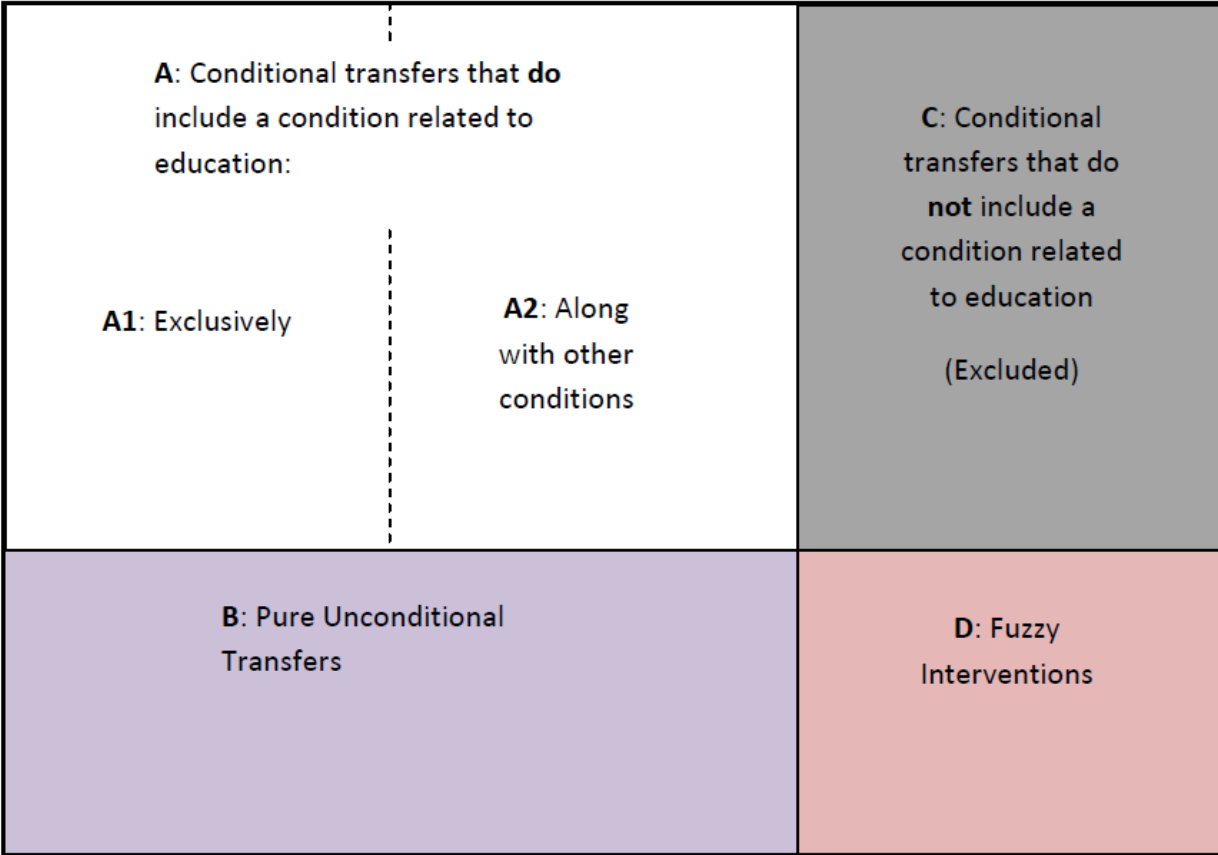
Table 10: Summary of Findings (attendance and test scores)

Panel A: Attendance	Odds of Child Being Enrolled in School:	Statistically Significant?*	# Effect Sizes*	Comments
Overall (vs. Control)	59% higher	Yes	20	A smaller number of studies assess the affect of CCTs and UCTs on attendance compared to enrollment. Both CCTs and UCTs have a significant effect on attendance. While the effect size is always positive, we do not detect significant differences between CCTs and UCTs on attendance.
UCT (vs. Control)	42% higher	Yes	5	
CCT (vs. Control)	64% higher	Yes	15	
CCT vs. UCT (regression)	17% higher	No	20	
Intensity of Conditionality (regression)	Increases by 8% for each unit increase in intensity of condition.	No	20	

Panel B: Test Scores	Standard Deviation Increase in Test Scores	Statistically Significant?*	# Effect Sizes*	Comments
Overall (vs. Control)	0.06	Yes	8	There are very few studies that analyze test scores. We have a total of 8 effect sizes measured from 5 studies. CCTs significantly increase test scores, though the size is very small at 0.08 standard deviations. We find no impact of UCTs on test scores. Additional research on the impact of CCTs and UCTs on test scores is needed. In order to include these results in meta-analysis tests should be conducted with the entire sample, and results presented in terms of standard deviations.
UCT (vs. Control)	0.04	No	3	
CCT (vs. Control)	0.08	No	5	
CCT vs. UCT (regression)	0.05	No	8	
Intensity of Conditionality (regression)	Increase of 0.02 standard deviations for each unit increase in intensity of conditions	No	8	

Notes: We consider a study to be statistically significant if it is significant at the 95% level or higher. We use the term effect size here instead of study since the studies that directly compare CCTs and UCTs have two effect sizes in the analysis. All other studies have one.

Figure 1: Cash Transfers to Households: Simple Taxonomy for the Purpose of Systematic Review



Notes: Included in the scope of the review: Sets A, B and D; Excluded: Set C

Figure 2: Number of Included Reports by Year of Publication

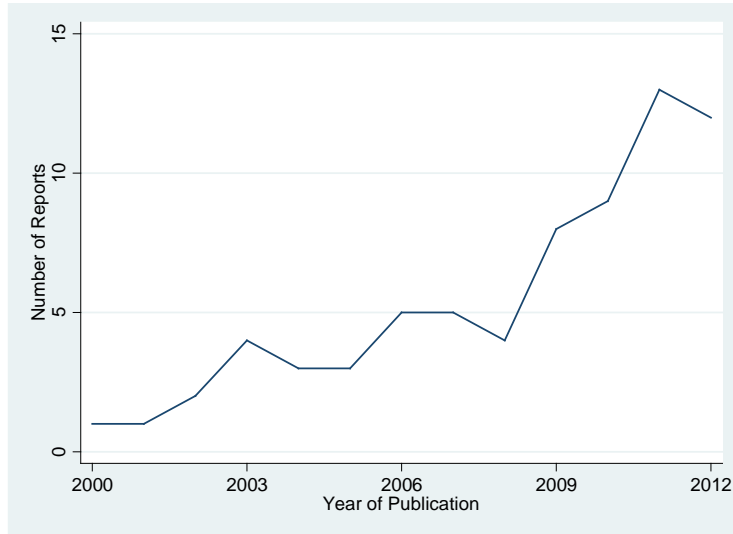


Figure 3: Impact of UCTs and CCTs on Enrollment

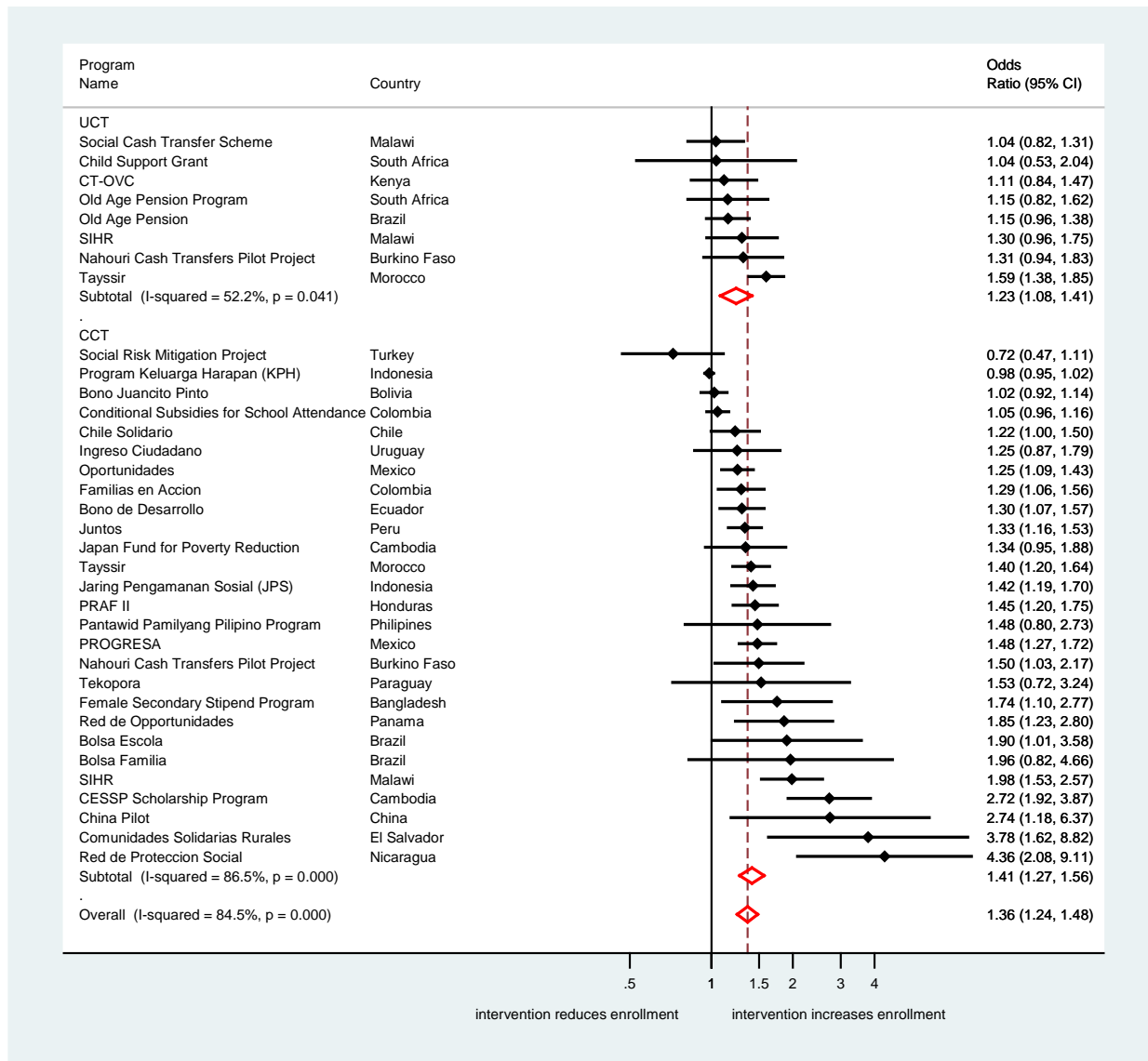


Figure 4: Impact of Intensity of Conditionality on Enforcement (0=None, 6=Enforcement on Attendance)

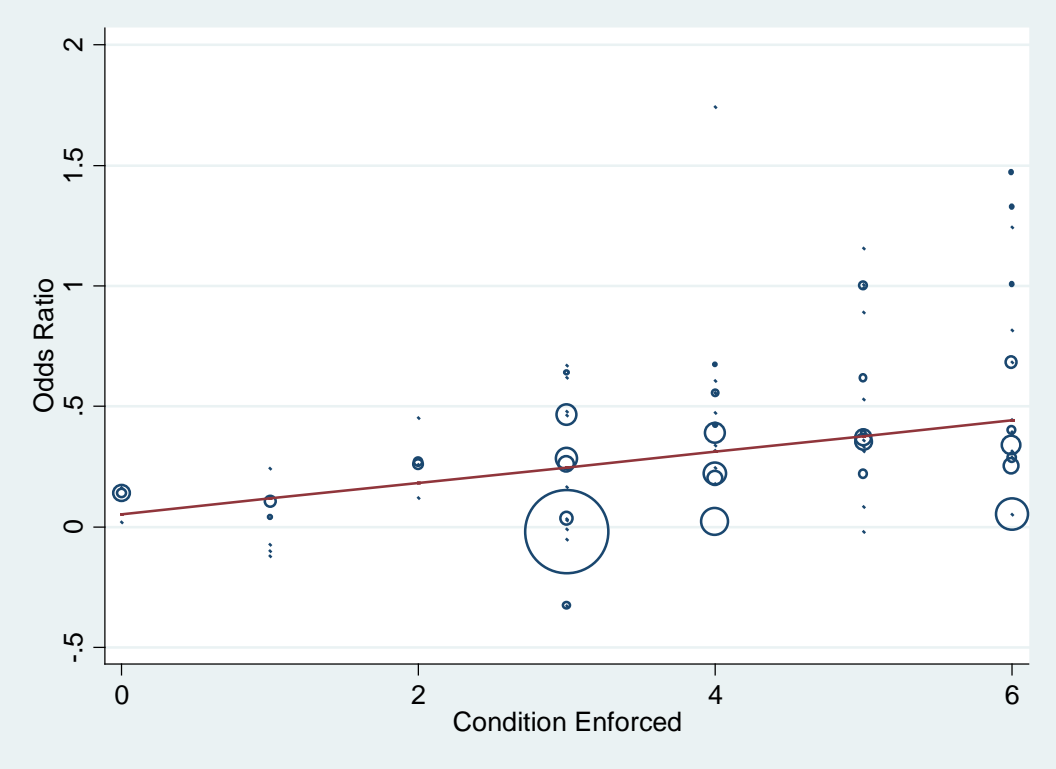


Figure 5: Impact on Enrollment by Group
(No Schooling Conditions, Conditions but no Enforcement, Conditions Enforced)

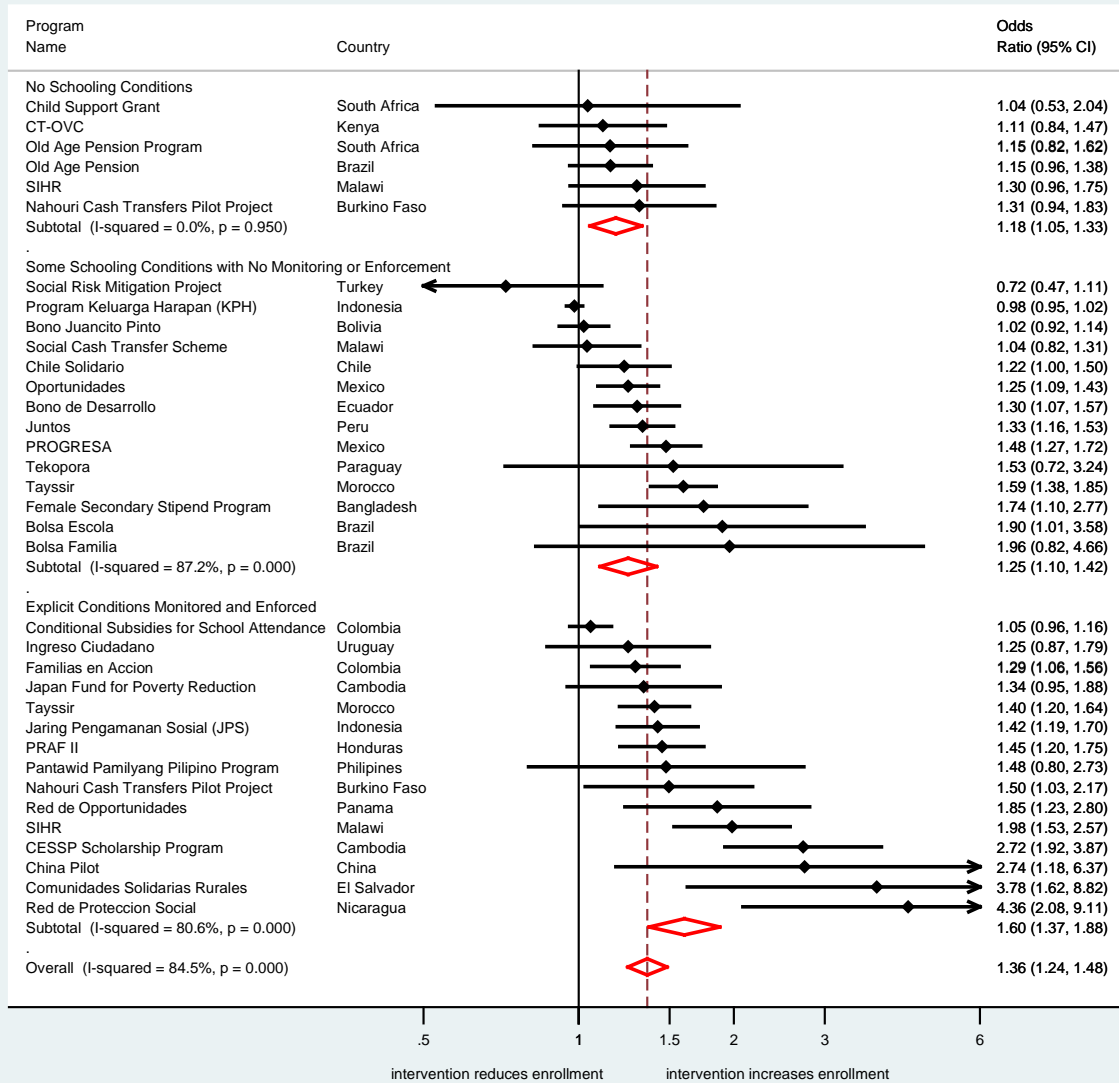


Figure 6: Impact of UCTs and CCTs on Enrollment (Boys)

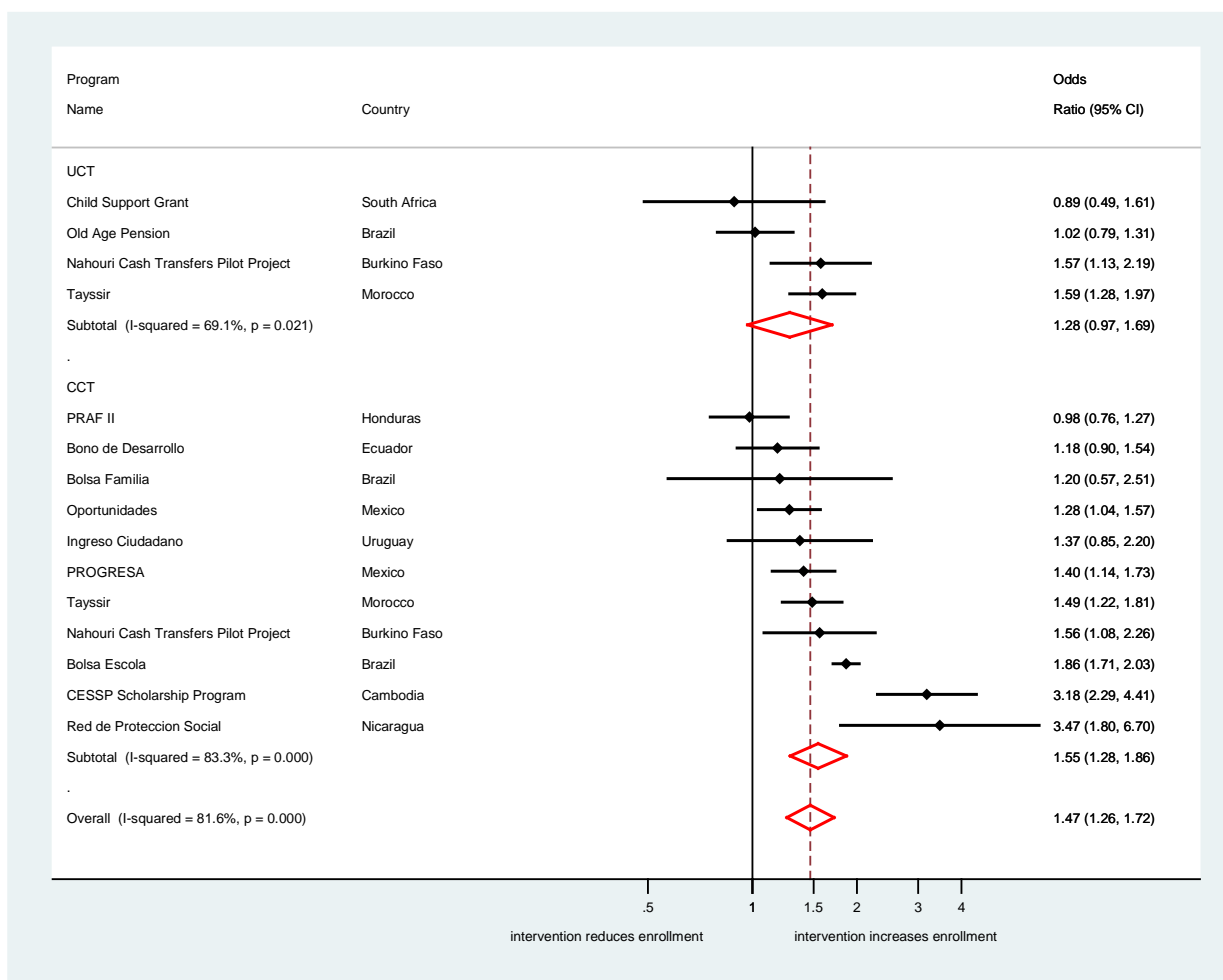


Figure 7: Impact of UCTs and CCTs on Enrollment (Girls)

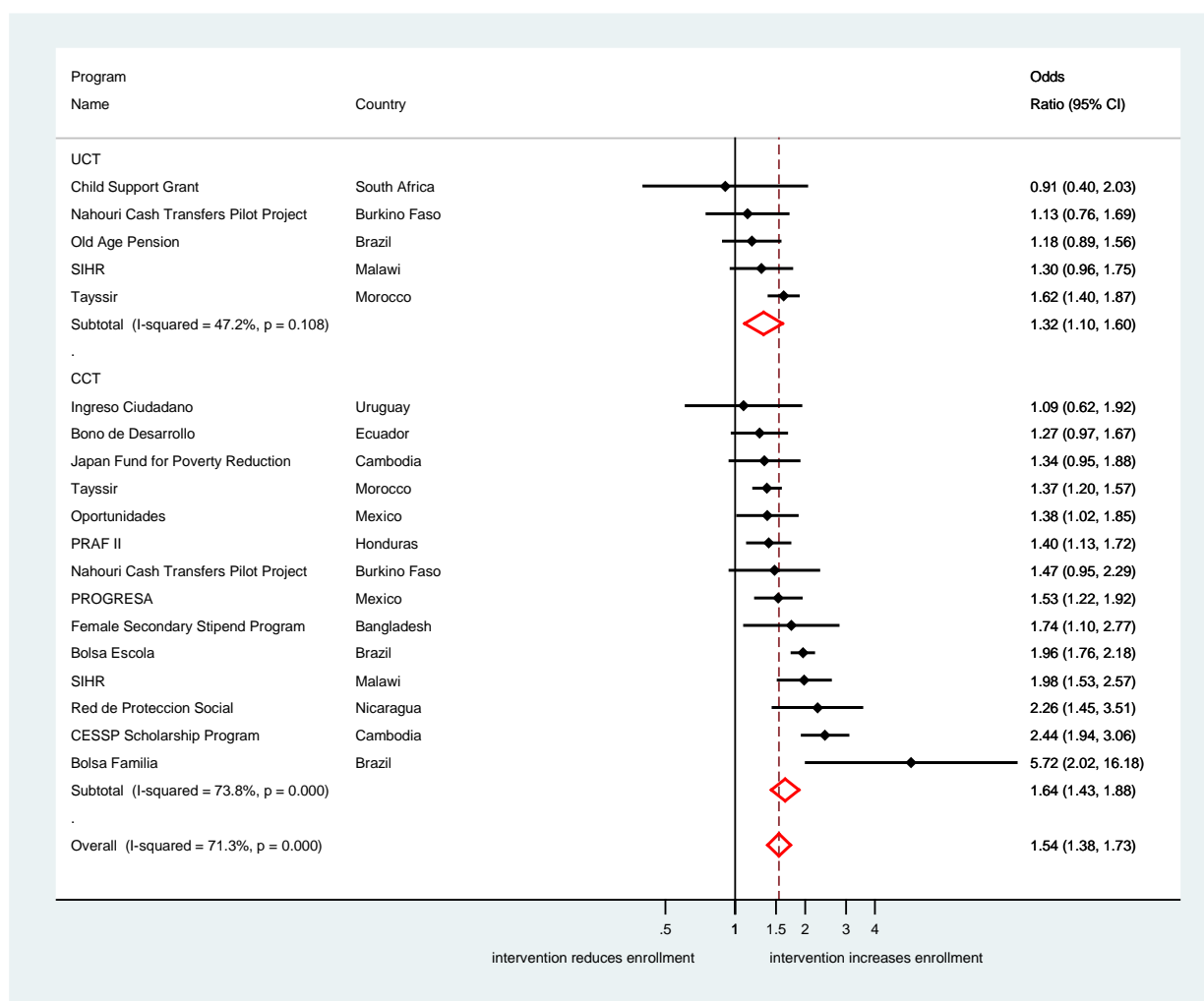


Figure 8: Impact of UCTs and CCTs on Enrollment (Primary)

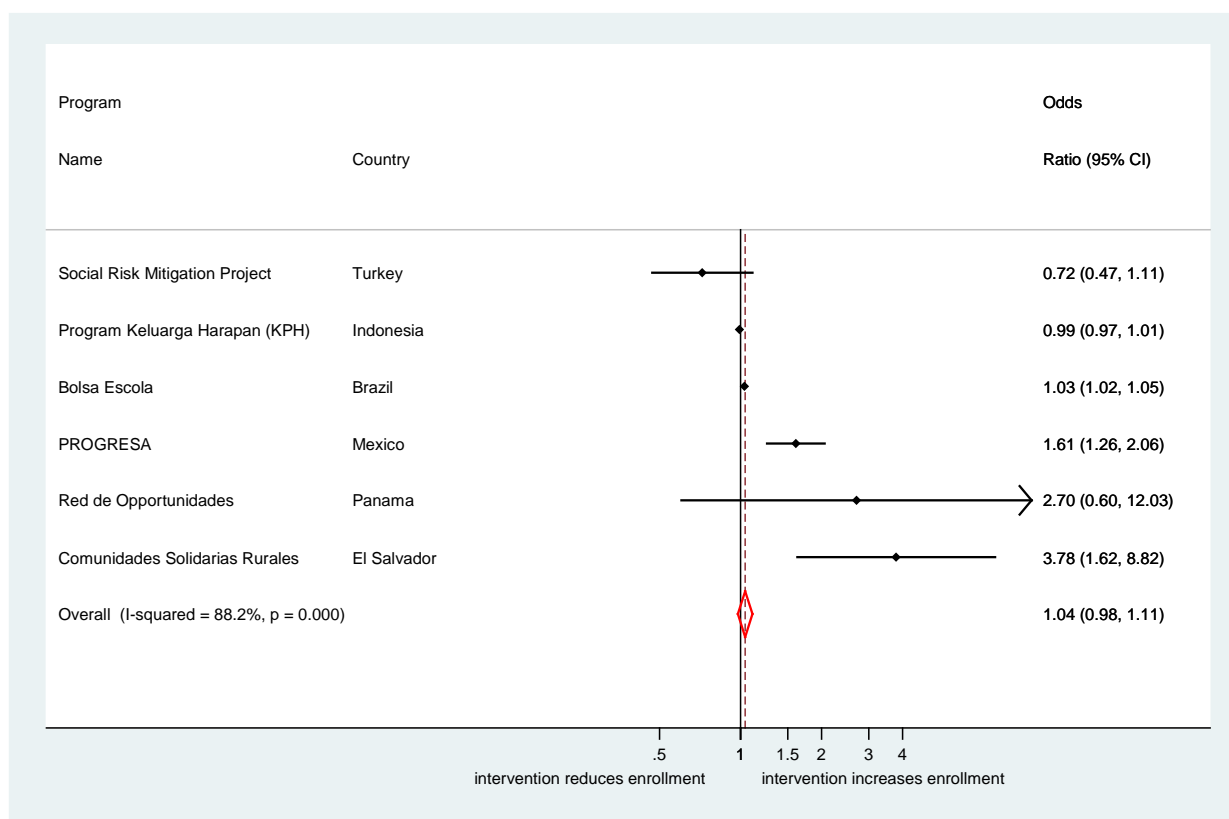


Figure 9: Impact of UCTs and CCTs on Enrollment (Secondary)

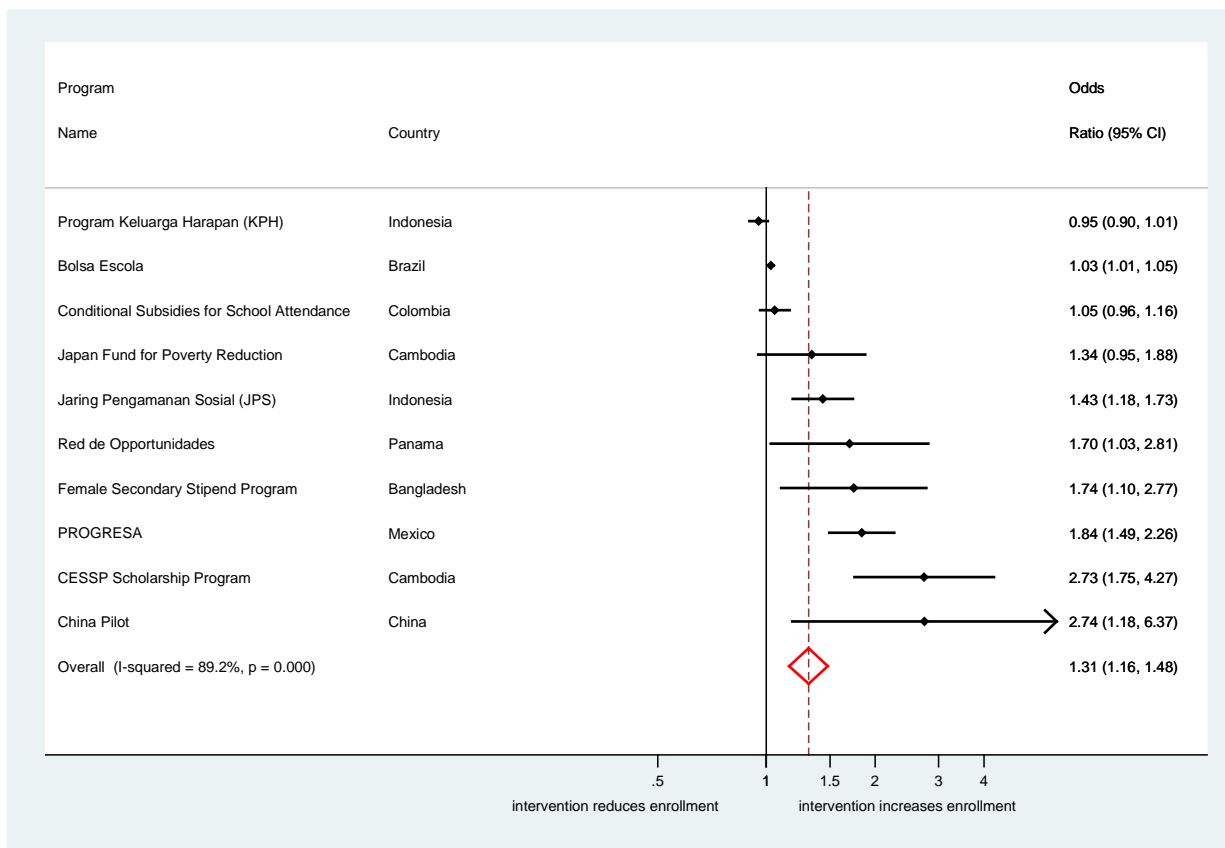


Figure 10: Impact of UCTs and CCTs on Attendance

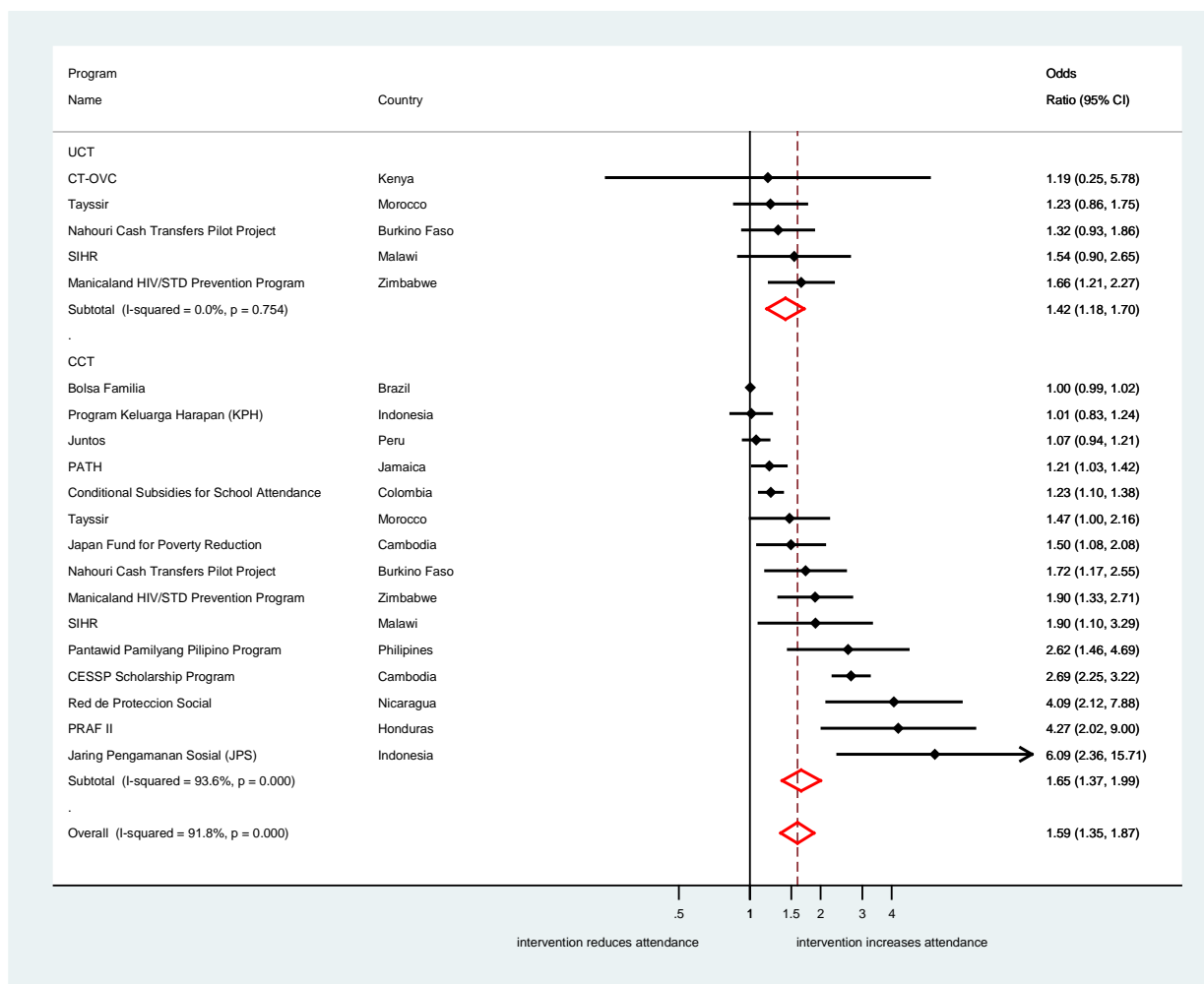


Figure 11: Impact of UCTs and CCTs on Test Scores

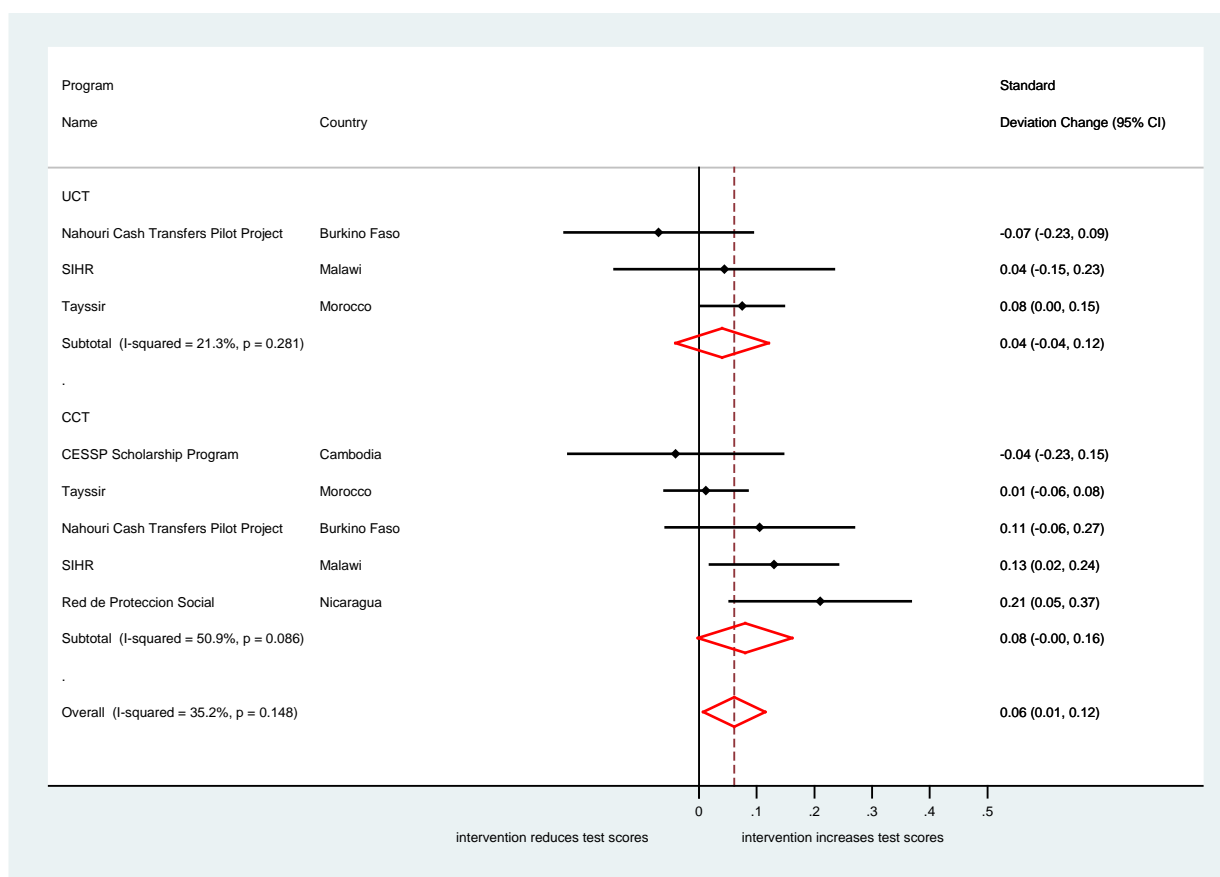


Figure 12: Funnel Plot for CCTs (Enrollment)

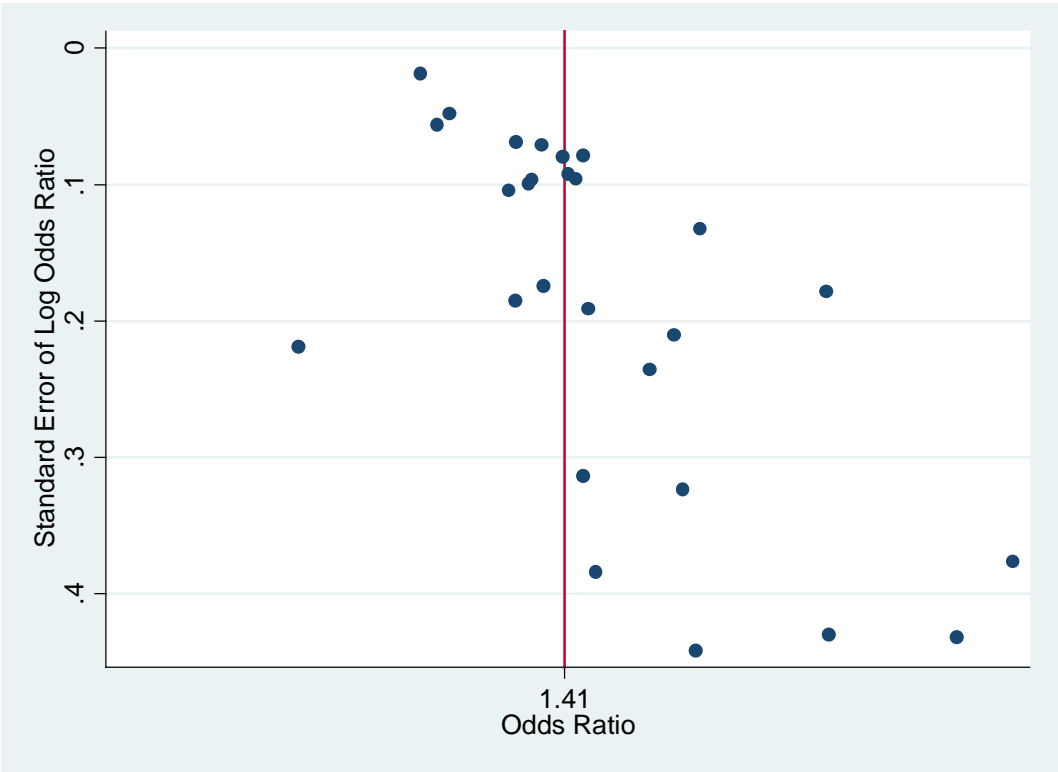
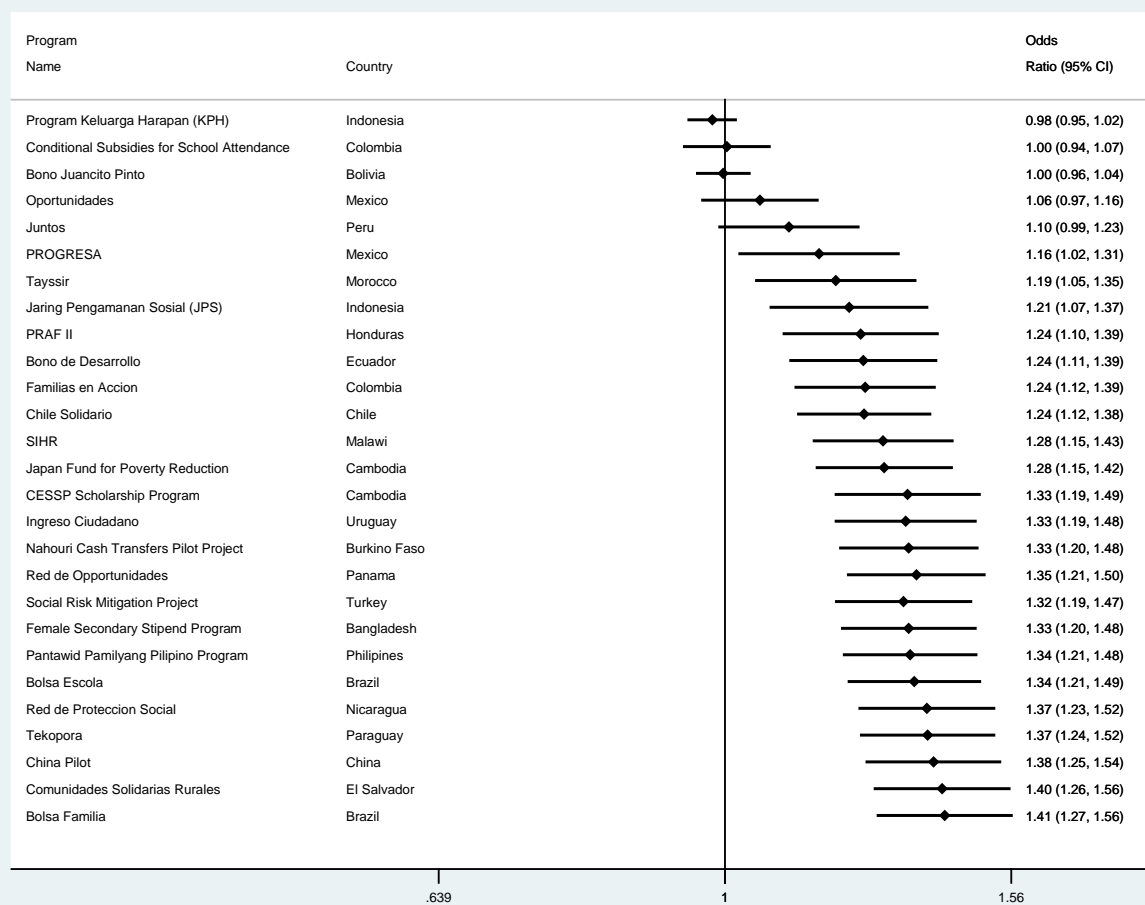


Figure 13: Cumulative Random Effects Meta-Analysis (Enrollment in CCT)



References

Adato, M. & Roopnaraine, T. 2010, "Women's Status, Gender Relations, and Conditional Cash Transfers", in Adato, M. & Hoddinott, J., eds. 2010, *Conditional Cash Transfers in Latin America*, Baltimore: Johns Hopkins University Press; Washington, D.C.: International Food Policy Research Institute.

Baird, S. Ferreira, F.H., Özler, B., & Woolcock, M. "Relative effectiveness and cost-effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: a systematic review", *Campbell Systematic Reviews*, 2013:8, DOI: 10.4073/csr.2013.8.

Borenstein, M., Hedges, L.V., Higgins, J.P.T. & Rothstein, H.R., 2009, *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd

Bourguignon, F., Ferreira, F. H. & Leite, P.G., 2003 "Conditional Cash Transfers, Schooling, and Child Labor: Micro-Simulating Brazil's Bolsa Escola Program," *World Bank Economic Review*, vol 17, pp. 229-254.

Bruns, B., Mingat, A. & Rakotomalala, R., 2003, *Achieving Universal Primary Education by 2015 - A Chance for Every Child*. Washington, D.C.: The World Bank

Fiszbein, A. & Schady, N. 2009, *Conditional Cash Transfers: Reducing Present and Future Poverty*, With Francisco H. G. Ferreira, Margaret Grosh, Nial Kelleher, Pedro Olinto, and Emmanuel Skoufias; Washington, D.C.: World Bank.

Gaarder, M. M., Glassman, A. & Todd, J., 2010, *Conditional Cash Transfers and Health: unpacking the causal chain*, *Journal of Development Effectiveness*, vol. 2, no 1, pp. 6-50.

Glewwe, P. & Kremer, M., 2006 "Schools, Teachers, and Education Outcomes in Developing Countries", in *Handbook of the Economics of Education*, eds. E. Hanushek & F. Welch, Elsevier, Volume 2, Pages 945-1017.

Grosh, M., Andrews, C., Quinta, R. & Rodriguez, C., 2011, "Assessing Safety Net Readiness in Response to Food Price Volatility," *Social Protection and Labor*, World Bank Discussion Paper No. 1118.

Hagen-Zanker, J., McCord, A., Holmes, R., Booker, F., & Molinari, E., 2011, "Systematic Review of the Impact of Employment Guarantee Schemes and Cash Transfers on the Poor." Overseas Development Institute (ODI) Systematic Review.

Hanlon, J., Barrientos, A. & Hulme, D., 2010, *Just Give Money to the Poor: The Development Revolution from the Global South*. Kumarian Press: Sterling, VA, USA.

Higgins JPT & Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Lagarde M, Haines, A., & Palmer N., 2007, "Conditional cash transfers for improving uptake of health interventions in low and middle income countries: a systematic review," JAMA, vol. 298, no. 16, pp.1900-1910.

Macours, K., Schady, N. & Vakis, R., 2008 "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment," World Bank Policy Research Working Paper No. 4759.

Özler, B. 2013, "Defining Conditional Cash Transfer Programs: An Unconditional Mess," Development Impact, May 13, 2013.
<http://blogs.worldbank.org/impactevaluations/defining-conditional-cash-transfer-programs-unconditional-mess>.

Paxson, C. & Schady, N., 2010, "Does money matter? The Effects of Cash Transfers on Child Development in Rural Ecuador," Economic Development and Cultural Change, vol. 59, no. 1, pp. 187–229.

Pritchett, L., Samji, S. & Hammer, J. 2012, "It's All About MeE: Using Structural Experimental Learning ('e') to Crawl the Design Space," UNU-Wider Working Paper No. 2012/104.

Saavedra, J.E. & Garcia, S., 2012, "Impacts of Conditional Cash Transfers on Educational Outcomes in Developing Countries: A Meta-analysis." RAND Corporation Working Papers, WR-921-1.

Tesliuc, E., del Nionno, C & Grosh, M. 2009, "Social Assistance Schemes Across the World Eligibility Conditions and Benefits," Center for International Policy Exchanges: University of Maryland School of Public Policy Conference Paper.

Todd, P.E. & Wolpin, K.I., 2006, "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," American Economic Review, vol 96, no. 5, pp. 1384–1417.

UNESCO, 2011, Institute for Statistics Fact Sheet No. 8, [online] Available at:
<http://www.uis.unesco.org/FactSheets/Documents/fs8-2011-en.pdf>.

Wilson, D. 2011, "Calculating Effect Sizes," [online] Available at:
http://www.campbellcollaboration.org/artman2/uploads/1/2_D_Wilson__Calculating_ES.pdf

World Bank, 2009, Education for All (EFA). [online] Available at:
<http://go.worldbank.org/I41DLBA8C0>

ⁱ However, as discussed later, some UCTs targeted to children may be framed in a way that suggests that the transfers are meant to support children's education or are accompanied by a discourse about the importance of education (Adato and Roopnaraine 2010). Benhassine et al. (2013) refer to such programs as 'labeled' rather than conditional or unconditional. Please see Figure 1 for a simple taxonomy of cash transfers that summarize these differences.

ⁱⁱ Baird et al. (2013) presents the full systematic review from which this paper is drawn. This systematic review can be found at: <http://www.campbellcollaboration.org/lib/project/218/>.

ⁱⁱⁱ By subgroup, we mean one of five options: overall, primary school, secondary school, boy, or girl.

^{iv} A complete list of included and excluded studies is listed in Appendix G.

^v The four experiments that have separate CCT and UCT arms are each counted as one study, but have two effect sizes, one for the CCT arm and one for the UCT arm.

^{vi} Given the range of design combinations that exist in our set of eligible studies, such an attempt is bound to be somewhat subjective. Below, we outline these seven categories as best as we can (providing examples) and list each study's assigned value in Appendix Table D1. We also discuss the robustness of our findings to alternative assignments. Finally, as our data are publicly available, interested readers can reconstruct such variables and assign them to different studies as they wish and rerun our meta-analysis.

^{vii} For details on why a specific study was excluded please contact the authors.

^{viii} There are two few studies to conduct this analysis among the UCT studies.

^{ix} A fourth such experiment from Zimbabwe does not report enrollment rates and has a moderate risk of bias.

^x There are a myriad of ways to design and implement a cash transfer program (see, for example, Pritchett, Samji, and Hammer 2012), not all of which will be observable to researchers or other policymakers. Hence, while evidence from systematic reviews such as this one (or from randomized experiments directly comparing CCTs and UCTs in some countries) can provide policymakers with a starting point, they are no substitute for careful consideration of all the variables that might moderate the effectiveness of such programs within a given setting.

^{xi} If not enough information to calculate an effect size was included in the report, it was ultimately excluded.

^{xii} The details relating to the coding of this variable are discussed later. The values of this variable assigned to each study are listed in Appendix Table D1.

^{xiii} Our search initially also included the following outcomes: cognitive tests, grade repetition, highest grade, grade completion and grade progression. However, the number of studies that reported these outcomes was small, with typically only one UCT study available for each of these outcomes.

^{xiv} Our advisory panel consists of Michelle Adato, Millennium Challenge Corporation; Nicholas Freeland, AusAID; Lisa Hannigan, AusAID; John Hoddinott, IFPRI, and Michael Samson, Economic and Policy Research Institute (South Africa).

^{xv} Note that this covariate adjusted success rate at follow-up in T is different than the raw success rate.

^{xvi} The standard deviation of the logistic distribution is equal to $\pi/\sqrt{3}=1.814$.

^{xvii} Standardizing test scores into comparable effect sizes requires mean test scores for each group, the pooled standard deviation (or the standard deviation and the sample size for each group).

^{xviii} We know that enrollment rates change over time in the absence of these programs, especially due to the age gradient in school enrollment. However, the study periods are usually short (1-2 years), so we prefer this approximation over either excluding the study from the review or assigning an ad hoc time trend to the control group.

^{xix} By subgroup, we mean one of five options: overall, primary school, secondary school, boy, or girl.

^{xx} This is the most conservative estimate we can make, meaning that the existence of multiple estimates for an outcome provides no improvement in precision, but only alter the ES. We view it to be a reasonable assumption. For the exact formulae we use, please see Chapter 24 in Borenstein et al. (2009), equations 24.4-24.5.

^{xxi} For the formulae used, please see Chapter 12 in Borenstein et al. (2009), equations 12.2-12.8.

^{xxii} Given the heterogeneity in the design and implementation of cash transfer programs around the world, the assumption of a random-effects model (that the true effect sizes come from a distribution) seems much more reasonable than that of a fixed-effects model (that there is one true effect size).

^{xxiii} To improve readability, we use a linear transformation of d (and its standard error) in our analysis, combined with the *eform* option for *metan*, which presents the pooled ES in Odds Ratios.

^{xxiv} Scopus includes a 100% search of Medline.