Cash Transfers and Health: Mechanisms and Impacts

David K. Evans — The World Bank
Brian Holtemeyer — IFPRI
Katrina Kosec* — IFPRI

November 5, 2015

Abstract

Why have conditional cash transfers only selectively improved health outcomes and investments in developing countries? We examine this question by exploiting the randomized introduction of a transfer program in Tanzania in 2010. We find several reasons for inconsistent results in the existing literature. First, health improvements took time to materialize; we find no improvements after 1.5 years of transfers, but significant improvements after 2.5 years, suggesting the importance of considering longer-term impacts. Second, health benefits were greatest in villages with more baseline health workers per capita, consistent with health improvements being sensitive to clinic capacity constraints. Finally, health did not improve due to more clinic visits, but instead due to more timely visits made when individuals are ill as opposed to when they have money, and these changes in attendance patterns are likely facilitated by significant uptake of health insurance. The results are robust to adjustments for multiple hypothesis testing.

JEL Classification: I18, I13, O12, H42
Keywords: Conditional cash transfers, service delivery, decentralization, health

*This study benefitted at various stages from experts at the World Bank, the International Food Policy Research Institute (IFPRI), the Tanzania Social Action Fund (TASAF) and elsewhere. At TASAF, the evaluation has been supported by the Executive Director Ladislaus Mwamanga, as well as the former Executive Director Servacius Likwelile. Amadeus Kamagenge has led TASAF input to the evaluation, and his entire team has contributed with substantive and logistical support. We are also grateful to Harold Alderman, Domenico Fanizza, Margaret Grosh, Melissa Hidrobo, Jef Leroy, David McKenzie, Berk Özler, Anna Popova, Dena Ringold, Shalini Roy, and Magnus Saxegaard for comments and helpful discussions. We received financial support from the CGIAR Research Program on Policies, Institutions, and Markets, the International Initiative for Impact Evaluation (3ie), the Strategic Impact Evaluation Fund (SIEF), and the Trust Fund for Environmentally and Socially Sustainable Development (TFESSD). Any comments or suggestions are welcome; the authors may be contacted at k.kosec@cgiar.org and devans2@worldbank.org.

1 Introduction

What role can cash transfers conditioned on health-seeking behavior play in alleviating the burden of poor health and limited access to formal medical care in Sub-Saharan Africa? Evidence from around the world suggests that conditional cash transfers (CCTs) can effectively alleviate extreme poverty and improve a range of human capital outcomes for children (Fiszbein and Schady, 2009; Independent Evaluation Group, 2011). In recent years, evidence from Africa has shown similarly positive results (Duflo, 2000, 2003; Baird et al., 2011, 2013; Akresh et al., 2014). As the evidence on cash transfers has grown, countries around the world have raced to adopt these programs. Almost every country in Latin America now has a CCT program (Fiszbein and Schady, 2009). Garcia and Moore (2012) report that, as of 2010, at least 35 countries in Sub-Saharan Africa had implemented some sort of cash transfer program, with 14 making transfers conditional on actions taken by the recipients.

While there is considerable evidence about the impacts of CCTs on education outcomes, less is known about their impacts on the full range of health-seeking and healthcare outcomes of the poor. This is unfortunate given that health itself affects school enrollment (Alderman et al., 2001) and learning productivity (Glewwe et al., 2001; Paxson and Schady, 2007). A 2014 global review found 142 studies showing the impact of cash transfers on education outcomes, but only 41 showing impacts on health and nutrition outcomes (Andrews et al., 2014). Furthermore, of 20 quantitative studies we identified examining the impact of cash transfers (conditional and unconditional) on health in Sub-Saharan Africa, the majority focus on a single outcome, such as clinic visits or child anthropometrics, and thus give only a partial understanding of their health impacts.

The existing literature on the health impacts of cash transfers yields mixed results; this suggests both the need for a better understanding of their full range of health-related impacts and of the factors that may account for selective successes in low income settings. Among studies from Africa, cash transfers have been shown to increase preventative health clinic visits for children (Akresh et al., 2014), improve physical and mental health (Baird et al., 2013), improve maternal healthcare utilitization for some mothers (Handa et al., 2015), and increase anthropometrics for girls—albeit not for boys (Duflo, 2003, 2000).[1] In contrast, anthropometric and nutritional impacts in Latin American studies have been very mixed, with null impacts in some cases (Morris et al., 2004; Paxson and Schady, 2010; Macours et al., 2008) and positive impacts in others (Fiszbein and Schady, 2009). There are similar cases of null impacts on health outcomes in Africa now emerging (Robertson et al., 2013; Aker, 2013; Haushofer and Shapiro, 2013). A global review of CCTs found significant positive impacts on child anthropometry (Leroy et al., 2009) while another found small, insignificant impacts (Manley et al., 2013).[2] This mix of promising and insignificant impacts, combined with the tendency of papers to focus on a single outcome, leaves us with little understanding of how transfers effectively improve health when they are successful. This paper seeks to fill that void.

Low income countries face serious health problems which cash transfer programs conditioned on health-seeking behavior might effectively address, but managing such programs is logistically challenging. Nowhere is the global burden of disease greater than in Sub-Saharan Africa. Life expectancy has increased by 20 years globally since 1970, but by only 10 years in Sub-Saharan Africa (Institute for Health Metrics and Evaluation, 2013; World Bank, 2013). While many countries in the region have witnessed a period of sustained economic growth over the last one to two decades, this growth has not proportionately helped the poor, who bear the greatest health burdens (Arbache and Page, 2007). Indeed, Africa's growth has a lower growth elasticity of poverty than any other region (World Bank, 2013). The region's

[1] In Burkina Faso and Malawi, some recipients received CCTs and others received unconditional cash transfers (UCTs). In Burkina Faso, UCTs did not increase health clinic visits (Akresh et al., 2014). In Malawi, both improved mental health, although the benefits were lower for CCTs of high monetary value, perhaps because the transfers then make up a significant proportion of the household budget, increasing the stress associated with complying with conditions (Baird et al., 2013). Early results from a UCT program in Kenya likewise showed no impacts on health outcomes (Haushofer and Shapiro, 2013).

[2] A few programs have examined more specialized cash transfer programs, linked specifically to maternal health investments or sexual behavior. Interventions in India and Nepal provided incentives for maternity services, with mixed results (Powell-Jackson et al., 2015; Powell-Jackson and Hanson, 2012). Interventions in Tanzania and Lesotho have provided incentives to remain free of sexually transmitted diseases, with positive outcomes (Bjorkman Nyqvist et al., 2015; De Walque et al., 2014).

health problems are in part the result of a pervasive lack of investment by the public or private sectors in health. This has contributed to a shortage of doctors and health facilities that make meeting health needs especially challenging. Additionally, to become embedded in a country's social protection system, CCTs require a strong central administrative structure—something uncommon in low income countries with extremely vulnerable populations. The effectiveness of CCTs also depends on low levels of corruption and little leakage of funds—pervasive problems in many African countries. These institutional and governance challenges raise the question of whether central government administration of a cash transfer scheme conditioning on health can be viable.

This paper considers a novel, community-managed CCT program launched in Tanzania in 2010 that circumvents some of the institutional challenges confronting implementation of a CCT conditioned on health-seeking behavior in Africa. In most CCT programs in Latin America (e.g., Brazil and Mexico), beneficiaries are identified via a centralized, means-tested registry of vulnerable households; likewise, payments are centralized. The Tanzanian community-managed CCT program modality relies on communities to identify the most vulnerable households, to verify compliance with conditions, and to distribute payments. As such, this model could potentially be replicated in many environments lacking a strong central governance structure.

In this context, we go beyond the more narrow analysis of one or two health outcomes to provide a detailed examination of the impacts of Tanzania's CCT on a full range of health investments, healthcare financing methods, the perceived and actual quality of healthcare facilities, as well as outcomes. [3] We take advantage of random assignment of the CCT to a sub-set of 40 villages among a larger group of 80 study villages to assess its causal impacts. Households were surveyed at baseline in 2009, again in 2011 after 18-21 months (about 1.5 years) of transfers, and finally in 2012 after 31-34 months (about 2.5 years) of transfers.

Our findings suggest several reasons for inconsistent health impacts of previous cash

[3] Impacts of this Tanzanian CCT program in other areas can be found in Evans et al. (2014).

transfer programs. First, health improvements take time to materialize. After 1.5 years in the program, there were no significant impacts on health outcomes. But 2.5 years into the program, transfers significantly reduced the likelihood of household members having experienced sick days in the last four weeks. These effects were greatest for young children. This suggests the importance of considering longer-term impacts. Second, health benefits were greatest in villages with more baseline health workers per capita, consistent with health improvements being sensitive to clinic capacity constraints. Finally, health does not improve due to more clinic visits, but instead due to more individual investments in preventative health measures and more timely visits made when individuals are ill as opposed to when they have money. Children in beneficiary households were about 50 percent more likely to own shoes. Also, while an initial increase in total clinic visits does not endure, beneficiaries were substantially more likely to seek treatment when ill, suggesting that clinic visits they did make were more appropriately timed to coincide with illness rather than with periods in which they could afford a visit. Existing literature shows that more timely clinic attendance in the face of illness improves child health outcomes in Tanzania (Adhvaryu and Nyshadham, 2015).

These changes in attendance patterns coincide with significant uptake of health insurance. Treatment led to higher expenditures on insurance, higher enrollment in the government-run health insurance program (the Community Health Fund), and a higher likelihood of financing treatment using health insurance as opposed to cash.[4] These results indicate that households greatly value insurance against health shocks. These shocks are relatively frequent; at baseline, a full 55 percent of households reported experiencing a health shock in the last five years.[5] Transfers had a short-term, positive impact on perceptions of the quality of health facilities—though those effects attenuated after an additional year of treatment, likely due to

[4]Little previous work has examined the impact of cash transfers on participation in health insurance programs. Evidence from Mexico suggests that participation in a CCT increased participants' awareness that they were enrolled in a health insurance program, but in that case, actual enrollment was automatic upon enrollment in the cash transfer program (Biosca and Brown, 2014).

[5]This includes a chronic or severe illness or accident of a household member, or a death in one's family.

habituation. Furthermore, we find no evidence that the quality of health facilities actually increased. There are no measurable impacts on mortality over the time frame of the project. Overall, this evidence suggests that cash transfers may help to lift the burden of disease in Africa, particularly when sufficient clinic staff and health financing instruments are available to complement the transfers. To test the robustness of the estimates, we carry out three different adjustments for multiple hypothesis testing and demonstrate that our main findings are robust.

The remainder of the paper is organized as follows: Section 2 provides background information on health and the healthcare system in Tanzania, as well as the health-related goals of Tanzania's pilot CCT program. Section 3 describes the evaluation design, data, and outcomes of interest. Section 4 describes our empirical specification and the groups over which we examine heterogeneous treatment effects and also presents balance tables showing the outcome of our randomization. Section 5 presents our main empirical results. Finally, Section 6 concludes.

2 Background

2.1 Healthcare and health in rural Tanzania

Tanzania is, in many respects, close to the Africa regional average in terms of health statistics. In 2012, the World Health Organization reported 17,318 malaria cases (per 100,000 population) in Tanzania versus 18,579 for Africa as a whole. Likewise, prevalence of HIV is at 3,082 (per 100,000) versus 2,774 for the region. Life expectancy at birth is 61 years versus 58 for Africa as a whole. The distribution of years of life lost across communicable diseases, non-communicable diseases, and injuries is very similar. Yet on some measures, Tanzania diverges significantly from the rest of the region. Its under-five mortality rate (54 per 1,000 live births) is just over half that of the region (95). Its maternal mortality ratio is almost twenty percent lower than the region as a whole. The health workforce, however, is weaker in Tanzania, with just 0.1 doctors per 10,000 population (versus 2.6 for Africa on average) and 2.4 nurses and midwives (versus 12.0 for Africa) (World Health Organization, 2014).

Crude measures of healthcare utilization, an area where cash transfers are likely to have the largest impact, suggest significant room for improvement, with only 34 percent contraceptive prevalence, 43 percent of pregnant women making a full set of antenatal care visits, and about half of births attended by skilled birth attendants (World Health Organization, 2014). Despite the limited health workforce, recent evidence disentangles the effect of using formal public health facilities from self-selection to demonstrate significant improvements in health outcomes for children who take advantage of these facilities (Adhvaryu and Nyshadham, 2015). In order to facilitate more consistent use of these facilities, the Tanzanian government in 1995 introduced a health insurance program called the Community Health Fund (CHF). This program is a voluntary, district–level prepayment scheme. Members pay a fixed annual fee of between 5,000 and 10,000 Tanzanian shillings (between $3 and $6 US), depending on the region, but then their entire family (up to 7 people) are exempt from any co-payments for visits to primary healthcare facilities (Marriott, 2011).[6] However, ten years after its introduction, the program had only an average enrollment rate of ten percent. At least two of the reasons cited for lack of participation were inability to pay or to see the rationale to insure (Kamuzora and Gilson, 2007). Insofar as liquidity has been a binding constraint, a cash transfer program might be expected to significantly impact participation.

2.2 Pilot CCT program

Tanzania's pilot CCT program began in January of 2010. It was implemented by the Tanzania Social Action Fund (TASAF, a social fund agency of the Tanzanian Government) and managed by local communities. Its principal goals were to increase investments in health for young children (ages 0 - 5) and the elderly (ages 60 and over) and to increase educational investments for children aged 7-15. It operated in three districts—Bagamoyo (70 km from

[6] Receipt of medications or tests may incur fees.

Dar es Salaam), Chamwino (500 km from Dar es Salaam), and Kibaha (35 km from Dar es Salaam), where 80 eligible study villages were randomized into treatment and control groups of 40 villages each, stratified on village size and district. The first payments were made in January 2010 and continued every 2 months thereafter. The amount of each transfer ranged from a US $12 minimum to a US $36 maximum, depending on household size and composition. These figures were based on the food poverty line; the CCT provided US $ 3 per month for orphans and vulnerable children up to 15 years of age (approximately 50 percent of the food poverty line) and US $ 6 per month for elderly of least 60 years of age (100 percent). In household survey data, the average reported payment was US $14.50.

In each village, a community management committee (CMC) comprised of 10-14 members was democratically elected by potential beneficiaries and was responsible for selecting beneficiaries and operating the program.[7] Each CMC received financial training and successfully managed at least one TASAF-supported project prior to the pilot. Immediately preceding the pilot, TASAF conducted an extensive communications and training program on the CCT at the regional, district, and village levels. CMCs were educated on how to identify and prioritize the poorest and most vulnerable households, and CMC members were then asked to carry out a survey of the poorest 50 percent of households. CMCs and community members understood that many fewer people than just those sampled would ultimately become beneficiaries of the program. They collected both objective data on households' poverty status and their own subjective rating of the household's poverty level (is the household exceptionally poor or not?).[8] TASAF then used the data to carry out a means test and propose to the community a ranking of households within that village by poverty level. CMCs then finalized and on occasion modified the beneficiaries list under the oversight of the

[7] Elections were held at a village meeting, under a closed ballot system.

[8] TASAF met with local leaders to discuss who they considered "vulnerable," and wished to target with the program. Following these discussions, TASAF provided broad guidelines to CMCs in all villages. Vulnerable children were defined as being abandoned or chronically ill, having one parent or both parents deceased, or having one or two chronically ill parents (e.g., with HIV/AIDS). Vulnerable elderly were defined as those with no caregivers, in poor health, or very poor. These guidelines helped the CMCs determine who should be interviewed as part of the census, and formed the basis for their subjective evaluations.

Village Council (VC) and with the endorsement of the Village Assembly (VA). On average, 23 percent of the villages' households were beneficiary households.

This oversight and validation helped promote community buy-in. Following beneficiary selection, CMCs in treatment villages continued to screen potential beneficiaries, communicate program conditions, transfer funds, and impose and enforce conditions. Control group households became beneficiaries almost three years after the treatment households, in November 2012. Random selection of the control and treatment villages was done after potential beneficiary households were identified in all 80 villages, to ensure comparability between vulnerable households identified in the treatment and control villages.

While CCT payments were made at the household level, conditions applied at the individual level. Children aged $0-5$ were required to visit a health clinic at least six times per year, elderly aged 60 and over were required to visit at least once per year, and no health conditions applied to other individuals. Children aged 7-15 were required to enroll in primary school and maintain an 80 percent attendance record. The CMC played a key role in monitoring conditions; they were responsible for collecting monitoring forms from health clinics and schools, updating records, delivering warnings when conditions were not met, making home visits to stay abreast of developments in beneficiary households, and conducting regular awareness sessions. A year and a half into the program, over 86 percent of beneficiary households reported that a member of the CMC had visited their household since the program began, and only 1.5 percent reported being asked for part of their transfer.

3 Evaluation Design and Data

3.1 Evaluation design

We evaluate the impact of the CCT program using three waves of data collected in all 80 program villages. A baseline survey was carried out during January - May 2009, and payments began in January 2010. A midline survey was conducted during July - September 2011 (18-

21 months, or about 1.5 years after transfers began) and an endline survey was conducted during August - October 2012 (31-34 months, or about 2.5 years after transfers began). The baseline survey included 1,764 households (a subset of beneficiary households) comprised of 6,918 individuals. The quantitative data collection was supplemented by two rounds of qualitative data collection employing focus group discussions and in-depth interviews.

3.2 Data and outcomes

In each of thee three survey rounds, we collected individual-level data on health clinic visits, ownership of protective footwear (shoes and slippers) by children, health (whether an individual was ill in the last four weeks, and for how many days in the last four weeks they were unable to perform their normal daily activities due to illness), ability to perform ordinary activities (doing vigorous activity, walking up hill, bending over or stooping, walking more than 1 km, walking more than 100 m, or using a bath or toilet), anthropometrics (height, weight, middle upper-arm circumference, and z-scores for height-for-age, weight-for-age, weight-for-height, and body mass index-for-age), and—for those who were sick—healthcare financing methods (free treatment, loans, cash or assets, or health insurance) and the location where medical care was sought (no care, public hospital, a public dispensary, a public health center, a private facility, a mission facility, a pharmacy or chemist, or a traditional healer).[9] Whether an individual was ill in the last four weeks captures the extensive margin of illness, while the number of days during that time that they were ill captures the intensive margin of illness. In each round we also collected household-level data on expenditure on formal insurance,[10] on the head's reported satisfaction with community healthcare facilities (whether they are poor, fair, good, or excellent), on whether the household participates in the government-run health insurance program known as the CHF, and on whether the head reports an inability

[9]Data on protective footwear were only collected for children aged $0-18$, data on anthropometrics were only collected for children aged $0-5$, and data on ordinary activities were only collected for elderly aged 60 and over. Sick individuals were asked to provide the primary financing method and health provider (in the case of multiple methods and providers) for their main health problem of the last four weeks.

[10]While we have data on expenditure on insurance, it is unfortunately not further disaggregated into expenditure on health insurance vs. other types.

to afford the CHF. We examine the program impacts on all of these dimensions of health and healthcare.[11]

Finally, at endline—though unfortunately not in previous rounds—we gathered village-level data on the main health facility, including its infrastructure quality as well as its staffing and treatment quality. Using principal components analysis (PCA) and taking the first principal component, we combined a number of variables into an infrastructure quality index,[12] and another set into a staffing and treatment quality index.[13] We additionally constructed an overall health facility quality index through a PCA that used the union of these two groups of health facility measures.[14] We then additionally examine program impacts on these three indices of health facility quality.

4 Methods and Empirical Strategy

4.1 Empirical specification

We carried out a baseline and two follow-up surveys, in 2011 and in 2012, to capture both short-term (1.5 years) and longer-term (2.5 years) impacts of this CCT. As assignment to treatment was random, we can recover causal estimates of the health impacts of the CCT on health-related outcomes by estimating the following empirical specification:

$$h_{it} = \beta_0 + \beta_1 2011_t + \beta_2 2012_t + \delta_1 T_i \times 2011_t + \delta_2 T_i \times 2012_t + \alpha_i + \epsilon_{it} \quad (1)$$

[11] Unfortunately, data on hygiene practices were not collected.

[12] These include the number of rooms to treat patients in the facility, the number of other rooms in the facility, the number of impatient beds in the facility, a dummy for whether the CHF operates in the village, and dummies for whether the facility has: a laboratory, bed for maternal deliveries, operating room, piped water, motorized vehicle, an incinerator, refrigerator, and electricity.

[13] These include the number of patients referred to another facility, the number of patients not able to be treated as much as desired, the number of patients seen in the past 30 days, the number of hours in the last week that a doctor was attending to patients, the number of doctors, the number of nurses, the number of assistants/staff, and dummies for whether the facility: provides HIV services beyond testing, has a compulsory vaccination program, has family planning services, has a childhood nutrition program, tests for HIV, tests for typhoid, tests for malaria, tests for anemia, tests for worms, tests blood sugar, tests blood pressure, and tests for typhoid.

[14] If a village has more than one facility, then data for the primary facility are used.

where i indexes individuals and t indexes the survey round. h_{it} is a health outcome, α_i are individual fixed effects, T_i=1 in the CCT treatment group and zero otherwise, 2011_t=1 at the time of the midline survey (July - September 2011) and zero otherwise, and $2012_t = 1$ at the time of the endline (August - October 2012) and zero otherwise. When we consider a household- or a village-level outcome, i instead indexes households or villages, respectively.

While treatment was rolled out at the village level, a number of households in treatment villages did not receive treatment—likely due to last minute changes in community prioritization or household refusal. Also, a small number of households in control villages received treatment—likely due to their close proximity to a treatment village. As a result, all reported results examine the effect of treatment on the treated by using assignment to treatment as an instrument for being treated; we use the fitted values from a first stage regression of treatment on assignment to treatment in place of T_i in equation 1.[15]

4.2 Heterogeneous treatment effects examined

We estimate the overall impacts of the CCT program as well as its impacts on several sub-groups. First, we examine impacts by age group. Given that health conditions applied only to children aged 0 – 5 and elderly aged 60 and over, and given that each of these two age groups has a different set of health issues and faced different conditions under the CCT, it is instructive to examine program impacts on them separately. Overall impacts include all individuals in the surveyed households, not only all individuals in the two sub-groups.

Second, for two central outcomes heavily linked to the supply side of healthcare—health clinic visits and illness—we examine heterogeneous impacts of the CCT program by baseline (2009) health clinic staff per capita. Specifically, we divide villages into two types: those with above-median and below-median healthcare staff per capita, where staff included doctors, nurses, and all other medical assistants and staff. This provides insight into how a village's initial human resources capacity to address healthcare needs and absorb added demand might

[15] We use Stata code written by Schaffer (2010).

mediate the impacts of a cash transfer program conditioned on health clinic visits. Of course, such results come with the caveat that healthcare personnel per capita may serve as a proxy for overall poverty or remoteness of the village—which could influence the interpretation of the results. In general, however, one would expect that villages with greater staffing levels would be better prepared to absorb potentially higher demand for healthcare, and therefore may see greater improvements due to the CCT program.

Finally, for outcomes heavily linked to the demand side of healthcare—whether one has government health insurance through the CHF, whether one treats illness when it strikes and where (public or private facilities), and how one finances treatment received for illness—we examine heterogeneous impacts of the CCT program by baseline (2009) household poverty, as captured by asset wealth. While all beneficiary households are poor—on average, only 23 percent of village households are beneficiaries—this allows us to observe whether it is the moderately poor (top half of beneficiaries in terms of asset wealth) or the extremely poor (bottom half of beneficiaries in terms of asset wealth) that benefit most from the program, and on which dimensions. In general, one would expect the extremely poor—being the most vulnerable—to see the most health benefits from a CCT program (Akresh et al., 2013). To capture asset wealth and divide households into two groups, we carried out a PCA using dummy variables for ownership of 13 assets.[16]

4.3 Outcome of the randomization

Despite randomization of villages into treatment and control, it is possible that some observable characteristics of treatment villages are significantly different than those of control villages. If this were the case, one would worry that treatment is correlated with observed and unobserved omitted variables. We address this concern in two ways. First, we show that randomization generally led to balance between treatment and control villages. Second, we use individual (or household- or village-, for outcomes that vary at those levels) fixed effects

[16]These include whether the household owns an iron, refrigerator, television, mattress or bed, radio, watch or clock, sewing machine, stove, bicycle, motorcycle, car or truck, wheelbarrow or cart, and mobile phone.

to account for any baseline imbalances.

A comparison of baseline sample means in treatment and control villages reveals balance on the vast majority of our outcomes (Table 1, Panel A). Across 43 outcomes, for only six are there significant differences at the 10 percent level between treatment and control, and for only three are there significant differences at the 5 percent level. Despite no overall differences in health between treatment and control villages, 0 - 5 year olds in treatment villages were ill or injured 0.06 percentage points (significant at the 10 percent level) more than the control group, and they had 0.51 more sick days every four weeks (significant at the 5 percent level). We also observe some imbalances in take-up of health-related products. 0 - 18 year olds in the treatment group were 10 percentage points less likely to own shoes than were those in the control group. Households in the treatment group also spent about 150 Tsh per year—about 11 cents[17]—more on health insurance than did control households (significant at the 10 percent level).

Treatment and control groups are quite balanced on other outcomes of interest. There were no significant baseline differences in the ordinary activities in which elderly individuals are able to engage. Similarly, among our seven anthropometric outcomes, treatment and control groups are only unbalanced on one: children in the treatment group had smaller weight-for-age z-scores (significant at the 10 percent level). Treatment and control groups are balanced on their healthcare provision locations and subjective rating of the quality of health facilities. Overall, household size is almost exactly the same, but treatment households had, on average, 0.13 fewer elderly members than did control households (substantively small but statistically significant at the 5 percent level). Finally, treatment and control groups were balanced on their methods of healthcare finance.

In Table 1, Panel B, we examine differences in baseline means between treatment and control groups for an array of demographic and housing characteristics. Individuals in the treatment and control groups are balanced on gender, education, age, the gender of the

[17] In 2009 the exchange rate ranged from 1,280 to 1,467 per U.S. dollar (Bank of Tanzania, 2015).

household head, and the presence of an improved roof, toilet facilities, and piped water. The only significant difference is that treatment households are less likely to have an improved floor (significant at the 5 percent level). To address imperfections in balance, we use individual (or household- or village-, for outcomes that vary at those levels) fixed effects.[18]

5 Results

We present impacts of Tanzania's CCT on a variety of health-related outcomes. First, we consider the impacts on health-seeking behavior, examining health clinic visits as well as investments that promote health or insure against health shocks, including purchases of shoes, slippers, and formal insurance. Second, we examine the impact on individuals' health and ability to perform ordinary activities. Third, we consider impacts on child anthropometrics. Fourth, we examine what individuals do when they are sick—including how they finance healthcare and where they go to receive it. In doing so, we also consider take-up rates of Tanzania's government-run health insurance program, the Community Health Fund (CHF). Finally, we examine impacts on subjective ratings of the quality of health facilities and compare those with impacts on objective measures of quality.

5.1 Health clinic visits

In Table 2, we document the impact of the CCT on the frequency of health clinic visits. We focus on overall impacts, impacts on children aged 0 - 5, and impacts on those over age 60, as these are the groups for whom health conditions applied. (Furthermore, at midline, data on clinic visits were only gathered for those two age groups.) At midline (1.5 years after treatment began), treatment was associated with 2.6 more visits per year for children aged 0 - 5 (column 3) and 1.1 more visits per year for those aged 60 and over (column 6).

[18] In the case of anthropometric outcomes, we use village × cohort fixed effects. These results are robust to instead using individual fixed effects, but since anthropometric measurements are only taken for children aged 0 - 5 and few children are in this age range for all of 2009 − 2012, we prefer the former specification.

These statistically significant effects, however, disappear at endline (2.5 years after treatment began) for both age groups. As the number of health visits is a count data outcome, we also show that our results are robust to instead estimating a Poisson model, in Table A3.[19]

Strikingly, the baseline mean for both age groups exceeds the number of clinic visits required by the program conditions. Children aged 0 - 5 already visited health clinics 8.3 times per year on average at baseline, compared to the the condition of 6 visits (59 percent of children already met this condition). Elderly aged 60 and older already visited health clinics 2.8 times per year on average at baseline, compared to the condition of 1 visit (65 percent of elderly already met this condition). Thus, the program's emphasis on clinic visits may have led households to initially increase visits even though the average household was already satisfying the condition. Subsequently—by the endline—households' understanding of the conditions may have improved, and they may have reduced visits to only those that were necessary, still exceeding the program conditions on average.

Despite differences in the point estimates, we cannot reject the null hypothesis that the effects of the CCT on health clinic visits in villages with few baseline health staff per capita (the bottom half of the distribution) are the same as those in villages with many health staff per capita (the top half). This is true overall and for both age groups. Given that health clinic staff per capita may reflect clinics' underlying ability to absorb increased demand, this is important; it suggests that the CCT's impact on clinic visits by beneficiaries in treatment villages is not being reduced by a lack of clinic staff. Of course, the possibility remains that the quality of visits is lower in poorly-staffed villages; we can only observe that a lack of clinic staff is not reducing the impact of the CCT on the number of clinic visits.

5.2 Health investments

While health clinic visits are an important aspect of individual investment in health, also important are purchases individuals make to prevent health problems from occurring or to

[19] Results hold whether we cluster standard errors at the village level or use bootstrapped standard errors.

cope with the risks posed by health shocks. Accordingly, we also examined whether the CCT impacted take-up of three health-related products: shoes, slippers (i.e., open-toed footwear) and insurance (Table 3). Already by midline, the CCT led to a significant, 19 percentage point increase in shoe ownership among 0 - 18 year old children that persisted until endline (column 1). A null impact on slipper ownership at midline changed to a significant, 9 percentage point increase by endline (column 2). This suggests that the CCT did not lead to a substitution of shoes for slippers, but rather increased take-up of both products by endline. Further, impacts were largest for ownership of shoes—which provide better protection. These impacts are remarkable considering baseline ownership rates of shoes and slippers were only 42 percent and 63 percent, respectively. We also find increased household expenditures on insurance; the CCT led to a 7 fold increase in expenditure on insurance by midline, and a 9 fold increase by endline (column 3).

The impact of the CCT on shoe and slipper ownership and on insurance expenditures are larger for the extremely poor than for the moderately poor in each of the two survey rounds (Table 3, Panel B). These differences are in several cases statistically significant. At midline, the extremely poor saw a significantly greater increase in shoe ownership and insurance expenditure than did the moderately poor, while at endline, the extremely poor had a significantly greater increase in slipper ownership than did the moderately poor. Overall, these results suggest that not only can a CCT increase take-up of products that tend to prevent health problems from occurring and help households cope with health-related risks, but also that the poorest of the poor are most likely to benefit.

5.3 Health and activities

Table 4 reports the effects of being in the treatment group on our two main health outcomes: whether or not an individual was ill or injured in the last four weeks, and the number of days that the individual was unable to perform their normal daily activities in the last four weeks due to illness (sick days) (Panel A). These capture, respectively, the extensive and intensive margins of illness. We see that at midline, treatment had no significant impact on either health outcome. However, at endline, treatment significantly reduced both the extensive and intensive margins of illness. These impacts were non-trivial: by endline, treatment resulted in a 4.6 percentage point reduction in the incidence of illness or injury in the last four weeks (a 17 percent decrease in the baseline mean incidence of 27.6 percent) and a half day decrease in sick days in the last four weeks (a 28 percent decrease in the baseline mean number of sick days, which was 1.64). These treatment effects were largest among young children (ages 0 - 5), for whom the reduction in incidence of illness is 11.8 percentage points and the reduction in sick days is 0.84. Similar results hold when we instead estimate a Poisson model that takes into account that this dependent variable is a count variable (Table A3). This suggests that the health benefits of the CCT took time to materialize, but were quite substantial, and were particularly marked for children.[20] We in fact find no significant overall program impacts for those aged 60 and over, either on the extensive or the intensive margins of illness.

We further explored the health impacts of the CCT by examining heterogeneous impacts by village health clinic staff per capita at baseline (Table 4, Panel B). Here, we find that the reductions in illness are driven by villages with more health staff per capita, with no significant impacts on health for villages in the bottom half of health staff per capita. For individuals in villages that were initially highly-staffed, the average reduction in incidence of illness across all age groups is 6.1 percentage points, and the reduction in sick days is 1. In the case of the reduction in sick days, the difference between the effect of the CCT on illness in poorly-staffed vs. well-staffed villages is statistically significant at the 5 percent level.[21] This suggests that the realization of health impacts may in fact be conditional on a village having sufficient available infrastructure and staff. This overall finding is even larger for the elderly; those living in well-staffed villages see a 1.6 day reduction in sick days during the

[20] It does not appear to be the case that the health benefits of the CCT were influenced by program-induced changes in household composition. Specifically, we find no evidence that treatment affected rates of mortality—overall, among children aged 0 to 5, or among elderly people aged 60 and older (Appendix Table A1). It also had no effect on overall household size, the number of 0 to 5 year olds, or the number of elderly living in the household (Appendix Tables A2 and A3).

[21] This is not the case for the incidence of illness, however, where the difference is not statistically significant.

last week (a 57 percent decrease in the group's baseline mean), compared with a statistically insignificant 0.8 day reduction for elderly in poorly-staffed villages—a differences that is statistically significant at the 1 percent level.

Despite overall health improvements, the CCT did not change the ordinary activities that elderly individuals were able to perform, as shown in Table 5. Specifically, it did not have robust impacts on individuals' reported ability to do vigorous activity, walk uphill, bend over or stoop, walk more than 1km, or use a bath or toilet, nor did it affect a simple 0 - 6 index of these activities (ordinary activities index). One exception is the ability to walk more than 100m (a dummy that had a very high baseline mean of 0.96); there, we find a very small negative impact of the CCT that is statistically significant. Overall, however, it is clear that the CCT did not have systematic impacts on the types of activities that individuals could perform; rather, it changed the number of days that they could perform their activities.

The findings on improved health demonstrate the importance of taking care to evaluate health outcomes after an appropriate period of time, as advocated in general by King and Behrman (2009). Positive health impacts do not appear after 1.5 years of transfers, but rather only after 2.5 years of transfers have been received.

5.4 Anthropometrics

Table 6 reports the effects of being in the treatment group on a number of anthropometric outcomes for children aged 0 - 5: height-for-age, weight-for-age, weight-for-height, body mass index (BMI)-for-age, height, weight, and middle upper-arm circumference (MUAC) (columns 1 - 7, respectively). These regressions use village $\times$ 6-month age cohort fixed effects since very few children were in the 0 - 5 age range for multiple observations during 2009 – 2012. We see that at neither midline nor endline does being in the treatment group have a statistically significant impact on any of these outcomes. The lack of anthropometric effects is striking given the large number of metrics we consider; it contributes to a mixed literature on the impacts of CCTs on child anthropometrics (Fiszbein and Schady, 2009). This result is less surprising when considering other analysis from this program, showing that households did not use their transfers to increase food consumption (Evans et al., 2014). Further, while expenditures on children's protective footwear and insurance are likely to improve child health, it is possible that their health benefits take more time to materialize.

5.5 Health insurance and healthcare financing methods

One of the most striking and consistent impacts of the CCT, found for all sub-groups, is an increase in participation in Tanzania's government-run health insurance program, the Community Health Fund (CHF) (Table 7). While we lack baseline data on participation rates in the CHF, at baseline only 3 percent of individuals who were sick during the last four weeks and sought treatment reported using health insurance to fund that treatment. Our empirical analysis shows that by endline, the CCT made households in treatment villages 38 percentage points more likely to participate in the CHF and 43 percentage points less likely to report being unable to afford the CHF (Panel A). The effect on both outcomes is slightly larger (but not statistically significantly different) for extremely poor households—those among program beneficiaries who were in the bottom half of asset wealth (Panel B). Note that this does not merely reflect an income effect that boosted expenditures on all goods; food consumption did not rise. Rather, households used the transfers to insure against health shocks. This is consistent with our finding of increased insurance expenditures.

Our qualitative data collection suggested that village health clinic staff went to the place at which beneficiaries picked up their transfers in order to encourage them to sign up for the CHF while they "still felt rich." If the sick were poorer and less able to hand over transfers to clinic staff, one might worry that the CCT increased take-up of health insurance without actually affecting how healthcare of the sick was financed. We find no evidence that this is the case. In particular, Appendix Table A4 examines the healthcare financing methods of individuals who reported being ill sometime during the last four weeks and who treated the illness. We find that at midline, the CCT reduced payment for healthcare using cash or an asset by 19 percentage points, which is a sizeable 29 percent decrease from the baseline mean of 65 percent. This effect was sustained at endline, where we see a 20 percentage point decrease. Those no longer financing treatment with cash and assets began using health insurance; we see a 17 percentage point increase at midline in use of health insurance to finance treatment, which swelled to a 20 percentage point increase by endline (all over a baseline mean of only 2.7 percent). By endline, this increase in usage of health insurance was statistically significantly larger for the extremely poor than for the moderately poor; they saw a 35 percentage point increase compared to a 26 percentage point increase for the moderately poor (this difference is significant at 5 percent level). We find no overall significant changes in use of loans or assistance to finance treatment, though we see a 9 percentage point reduction in reliance on free treatment at endline. The CCT not only increased take-up of health insurance, but also increased its usage among the sick by leading them to finance treatment with health insurance rather than cash or assets.

5.6 Choice of healthcare provider type

Table 8 examines the impacts of the CCT on healthcare provider decisions of individuals who reported being ill sometime in the last four weeks.[22] These individuals either failed to treat their main health problem (15 percent did so at baseline) or visited one of several different types of public and private sector healthcare providers. We find that at midline, the CCT led to a significant reduction in failure to seek treatment (13 percentage points)—although this impact was not sustained at endline (column 1). We also find that at both midline and endline, the CCT led to a significant increase in use of public dispensaries (column 2). Use of public dispensaries increased by 19 percentage points at midline, and by 16 percentage points at endline. At the same time, we find no evidence that the CCT impacted use of private healthcare providers (columns 5 - 8). Rather than switching from the private to the public sector, individuals switched from a failure to treat their illness to treatment at a public

[22] At baseline, 28 percent of individuals reported being ill or injured in the last four weeks; 85 percent sought treatment for their illness—49 percent in the public sector and 36 percent in the private sector.

dispensary. Importantly, the CCT did not lead individuals to treat illness at a public hospital or a health center. These are larger facilities that would typically offer more services (and potentially more qualified healthcare staff), but which are usually further from (typically outside) an individual's village. We do not find significant differences in the impacts of the CCT on the type of healthcare provider visited by baseline household wealth.[23]

5.7 Subjective ratings of health facilities

Because the CCT required individuals to visit health clinics regularly, one might suspect that it provided greater information about the quality of those clinics. Given greater demand for healthcare and greater participation in the CHF, one might also expect clinics to make investments in their infrastructure, staff, and supplies that could lead to objectively higher quality facilities and care. This motivated us to examine two outcomes: household heads' subjective impressions of the quality of their village health facilities (specifically, whether they were poor, fair, good, or excellent in each of the three rounds of data collection) and actual facility quality (village level data, only available only at endline).

Table 9 shows that transfers led to significantly higher subjective assessments of the quality of health facilities at midline, though the effects disappeared by endline. Specifically, at midline, the treatment group saw a 14 percentage point higher share of individuals rating health facilities as good or excellent relative to the control group, and a 5 percentage point higher share of individuals rating health facilities as excellent. The finding that these differential subjective quality ratings disappeared by endline is likely due to habituation. It is also complemented by analysis showing that health facilities had not actually improved—either in terms of an index of infrastructure quality, in terms of an index of staffing and treatment quality, or in terms of an overall index of health facility quality—by endline (Table

[23] This analysis is attended by the caveat that the CCT program may have changed the composition—in terms of demographic and other characteristics—of those who report being sick sometime in the last four weeks. As the sick comprised only 29 percent of the sample at baseline, bounding these estimates by making different assumptions about the likely treatment choices of the well, had they been sick, would require making assumptions about a full 71 percent of our sample. As such, we have not attempted any bounding analysis.

10). An important caveat is that these indices of objective health facility quality may fail to capture other important aspects of quality, such as facility cleanliness, ease of getting an appointment, wait times, doctor quality and knowledge, or efficiency of care (such as whether treatment of a problem requires multiple visits or can be completed in a single visit). We unfortunately lack data on these aspects of quality. Overall, however, it does not appear to be the case that healthcare quality changed appreciably by endline.

5.8 Robustness: multiple hypothesis testing

One of the key contributions of this paper is that it provides a comprehensive treatment of the impacts of a CCT on health; rather than limiting ourselves to a few more standard outcomes, we have focused on a large variety while also carefully examining how impacts vary with beneficiary age, poverty level, and the initial human resources capacity of the healthcare system. As such, we take seriously the need to account for the risk of finding false positives when testing multiple hypotheses.[24] A growing literature recognizing this risk has advanced a number of potential correction methods.

Among the most widely-used are the Benjamini and Hochberg method (hereafter BH) and the Benjamini–Krieger–Yekutieli method (hereafter BKY) (Benjamini and Hochberg, 1995; Benjamini et al., 2006). Both methods control for the false discovery rate (FDR)—the expected share of rejections of a true null (i.e., type I errors). We implement both and compute the q-values (i.e., the p-values corrected for multiple hypothesis testing) emerging from each. As a third test, we apply the Bonferroni correction—a method of controlling the family-wise error rate (FWER), or the rate of making *any* type I error. The Bonferroni correction involves multiplying each p-value by the number of tests performed; while it has the advantage of being simple to compute, it suffers from poor power (Anderson, 2008) and is often used as an upper bound on the family-wise error rate (Hochberg, 1988). We accordingly rely primarily on the BH and BKY results, but we take Bonferroni corrections

[24] Correcting for multiple hypothesis testing in the policy evaluation literature has not kept pace with the frequency with which such corrections are made in other fields (Anderson, 2008).

as providing a useful guide to the lower bound of the significance of our results.

Performing the BH method of controlling the false discovery rate at some researcher-chosen, desired significance level q^* is very simple.[25] Benjamini and Hochberg (1995) describe the process for the case of m hypotheses:

> Consider testing $H_1, H_2, ..., H_m$ based on the corresponding p-values $P_1, P_2, ..., P_m$. Let $P_1 \leq P_2 \leq ... \leq P_m$ be the ordered p-values, and denote by H_i the null hypothesis corresponding to P_i. Define the following Bonferroni-type multiple-testing procedure:
>
> let k be the largest i for which $P_i \leq \frac{i}{m} q^*$
>
> then reject all H_i $i = 1, 2, ..., k$.

The BKY method of controlling the false discovery rate is a two-stage procedure that essentially repeats the BH procedure twice:

stage 1: Apply the BH procedure using $q' = \frac{q^*}{1+q^*}$ in the place of q^*. Let r be the number of rejected hypotheses. If $r = 0$, then do not reject any hypothesis and stop. If $r = m$, then reject all m hypotheses and stop. Otherwise, continue.

stage 2: Let $\hat{m}_0 = m - r$. Apply the BH procedure again using $q'' = q' \frac{m}{\hat{m}_0}$ in the place of q'.

Both BH and BKY simply tell one whether a hypothesis is rejected at the researcher's desired significance level, q^*. However, for each hypothesis being tested, it is valuable to know the *smallest* q^* for which the hypothesis can be rejected—i.e., the q-value of the test. To calculate these q-values, we repeatedly ran these tests for values of q^* between 0 and 1.[26]

Appendix Table A5 reports the resulting q-values from the three correction procedures described. We display q-values for each hypothesis that was statistically significant at the 10 percent level or higher in our main tables, to test whether it remains statistically significant at conventional levels after correcting for multiple hypothesis testing. For each correction procedure, a group of hypotheses is defined by the follow-up survey round (midline or endline) and broad type of outcome variable being tested (e.g., clinic visits, take-up of health-

[25] Özler (2014) has a very accessible blog post on this topic.

[26] Starting at 1, the q^* value is reduced by 0.0001 until the q-value reaches 0.0001. We appreciate Anderson providing this Stata code.

related products, child anthropometrics, or ordinary activities). This is usually equivalent to grouping together all of the hypotheses within a table for a given survey round.[27]

In total, we estimate 64 statistically significant (at the 10 percent level) impacts of the CCT on some variable, group of individuals, and survey round. Correcting for multiple hypothesis testing using the BKY method, 50 of these remain significant at conventional levels. When we instead use the BH method, 45 remain significant, and with the more conservative Bonferroni method, 38 remain significant. We take this as strong evidence that the CCT had significant impacts. Our main conclusions regarding the impact of the CCT still strongly hold. Under all three multiple testing correction methods, the CCT significantly increased health clinic visits at midline but did not affect them by endline. Furthermore, it significantly boosted childrens' shoe ownership as well as household expenditure on insurance at both midline and endline. Reductions in the intensive margin of illness by endline are still significant with both BKY and BH, even though they do not survive the highly conservative Bonferroni correction. However, reductions in the extensive margin of illness do not survive any of the three correction methods. This may indicate that the CCT's ability to reduce illness is principally concentrated in its reduction of the severity and intensity of illness—and generally how debilitating that illness becomes—rather than incidence of illness. While we found one odd result in which the CCT reduced the reported ability of the elderly to walk more than 100m, this finding does not survive any of the three corrections. Very notably, under all three corrections, findings that the CCT increases participation in and the affordability of the government-run health insurance program, the CHF, remain statistically significant at the 1 percent level. Similarly, findings that the CCT reduced the likelihood of failing to treat illness at midline, and increased visits to public dispensaries (but not private ones) at both midline and endline remain robustly significant under all three corrections.

[27] There are three exception to this. In Table table 3, the two individual-level outcomes (owning shoes and slippers) are grouped separately from insurance expenditures (household-level). In Table 4, the two outcome types—the extension margin of health and the intensive margin of health—are grouped separately. And in Table 3, the 6 specific activities of daily living are considered separately from the ordinary activities index; estimation of an index serves as an additional check on our corrections for multiple inference.

Finally, at both midline and endline, the finding that the CCT reduced use of cash or assets to finance healthcare and boosted usage of health insurance are again statistically significant under all three corrections. We conclude that the paper's main conclusions are strongly robust to corrections for multiple hypothesis testing.

6 Conclusion

Recently, it seems we are entering the era of cash transfers, with not only an uptick in programs and research but also in calls to expand cash transfers dramatically further, using them as a fair way to distribute the gains from natural resources as well as to alleviate poverty (Blattman, 2014; Blattman and Niehaus, 2014; Kenny, 2015; Devarajan et al., 2013). One concern in the face of such calls is whether the poor will use the resources poorly. A recent review of cash transfer programs around the world, both conditional and unconditional, shows that households do not disproportionately use cash transfer resources on items like alcohol and tobacco (Evans and Popova, 2014). But even so, one may wonder whether low-income, often poorly-educated household heads will allocate resources from cash transfers in ways that will provide returns.

This paper provides evidence that, in the context of a largely community-run conditional cash transfer (CCT) program in Tanzania, households receiving transfers are significantly less likely to lose working days as a result of illness. The statistically significant improvements in health outcomes after 2.5 years are particularly striking given that the CCTs' impact on the number of clinic visits was not sustained. If health improvements were not driven by increased visits, nor by increased consumption, then what was the cause? The evidence suggests that households used their transfers to invest in reducing the risk of high heath care costs. Households invested in footwear for their children, which reduces exposure to health risks. Households were substantially more likely to invest in the government-run health insurance program. They went on to utilize that health insurance to finance clinic visits when ill. Although the total number of clinic visits was not higher among beneficiary households, participation in the insurance program meant that those households could attend the clinic when they most needed, rather than letting immediate financial liquidity determine when, in the course of an illness, the household could visit the clinic. This is consistent with findings from Adhvaryu and Nyshadham (2015), who demonstrate that households that access formal sector malaria treatment in a more timely way have better health outcomes. The number of visits matters only in part; the timing of visits is also crucial, and the insurance program makes that timing more flexible. The availability of such health financing instruments may be an important consideration if countries desire to fully reap the health gains associated with cash transfers.

The health effects are particularly prominent in villages with more health workers per capita. In other words, cash transfers can most effectively improve health outcomes when the health infrastructure is of sufficient quality to meet the rising demand brought about by conditions. Clear evidence demonstrates that healthier households are likely to have higher incomes, which then drive better health in a virtuous cycle (Strauss and Thomas, 1998). The evidence from this study demonstrates that supply-side investments (in health infrastructure) combined with cash transfers (to permit households to insure against health shocks) may be critical catalysts to that virtuous cycle.

References

Adhvaryu, A. and A. Nyshadham (2015). Returns to treatment in the formal health care sector: Evidence from Tanzania. *American Economic Journal: Economic Policy 7*(3), 29–57.

Aker, J. C. (2013). Cash or coupons? testing the impacts of cash versus vouchers in the Democratic Republic of Congo. *Center for Global Development Working Paper*.

Akresh, R., D. De Walque, and H. Kazianga (2013). Cash transfers and child schooling: Evidence from a randomized evaluation of the role of conditionality. *World Bank Policy Research Working Paper* (6340).

Akresh, R., D. de Walque, and H. Kazianga (2014). *Alternative Cash Transfer Delivery Mechanisms: Impacts on Routine Preventative Health Clinic Visits in Burkina Faso*. University of Chicago Press. http://www.nber.org/chapters/c13377.

Alderman, H., J. R. Behrman, V. Lavy, and R. Menon (2001). Child health and school enrollment: A longitudinal analysis. *The Journal of Human Resources 36*(1), 185–205.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484).

Andrews, C., A. Lopez, and J. Baez (2014). What are we learning on safety net impacts? reviewing evidence from 2010-2013. *World Bank Social Protection Discussion Paper Series*.

Arbache, J. and J. Page (2007). *Patterns Of Long Term Growth In Sub-Saharan Africa*. The World Bank. http://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-4398.

Baird, S., J. De Hoop, and B. Özler (2013). Income shocks and adolescent mental health. *Journal of Human Resources 48*(2), 370–403.

Baird, S., C. McIntosh, and B. Özler (2011). Cash or condition? evidence from a cash transfer experiment. *The Quarterly Journal of Economics*, qjr032.

Bank of Tanzania (2015). Interbank foreign exchange market summaries. https://www.bot-tz.org/FinancialMarkets/IFEMsummaries/IFEMsummaries.asp.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika 93*(3), 491–507.

Biosca, O. and H. Brown (2014). Boosting health insurance coverage in developing countries: do conditional cash transfer programmes matter in Mexico? *Health policy and planning*, czt109.

Bjorkman Nyqvist, M., L. Corno, D. De Walque, and J. Svensson (2015). Using lotteries to incentivize safer sexual behavior: evidence from a randomized controlled trial on HIV prevention. *World Bank Policy Research Working Paper* (7215).

Blattman, C. (2014). Let them eat cash. *The New York Times*. http://www.nytimes.com/2014/06/30/opinion/let-them-eat-cash.html?_r̄0.

Blattman, C. and P. Niehaus (2014). Show them the money. *Foreign Affairs 93*(3), 117–126.

De Walque, D., W. Dow, and R. Nathan (2014). Rewarding safer sex: conditional cash transfers for HIV/STI prevention. *World Bank Policy Research Working Paper* (7099).

Devarajan, S., M. M. Giugale, H. Ehrhart, T. M. Le, and H. M. Nguyen (2013). The case for direct transfers of resource revenues in Africa. *Center for Global Development Working Paper* (333).

Duflo, E. (2000). Child health and household resources in South Africa: Evidence from the old age pension program. *American Economic Review*, 393–398.

Duflo, E. (2003). Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in South Africa. *The World Bank Economic Review 17*(1), 1–25.

Evans, D., S. Hausladen, K. Kosec, and N. Reese (2014). *Community-Based Conditional Cash Transfers in Tanzania: Results from a Randomized Trial*. World Bank Publications.

Evans, D. and A. Popova (2014). Cash transfers and temptation goods: a review of global evidence. *World Bank Policy Research Working Paper* (6886).

Fiszbein, A. and N. Schady (2009). *Conditional cash transfers: reducing present and future poverty*. World Bank Publications.

Garcia, M. and C. M. Moore (2012). *The cash dividend: The rise of cash transfer programs in Sub-Saharan Africa*. World Bank Publications.

Glewwe, P., H. G. Jacoby, and E. M. King (2001). Early childhood nutrition and academic achievement: A longitudinal analysis. *Journal of Public Economics 81*(3), 345–368.

Handa, S., A. Peterman, D. Seidenfeld, and G. Tembo (2015). Income transfers and maternal health: Evidence from a national randomized social cash transfer program in Zambia. *Health Economics*. http://dx.doi.org/10.1002/hec.3136.

Haushofer, J. and J. Shapiro (2013). Household response to income changes: Evidence from an unconditional cash transfer program in Kenya. *Massachusetts Institute of Technology*.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika 75*(4), 800–802.

Independent Evaluation Group (2011). Social safety nets: An evaluation of world bank support, 2000–2010. Washington, DC: Independent Evaluation Group, the World Bank Group.

Institute for Health Metrics and Evaluation (2013). The global burden of disease: Generating evidence, guiding policy — Sub–Saharan Africa regional edition. http://documents.worldbank.org/curated/en/2013/08/18187588/global-burden-disease-generating-evidence-guiding-policy-sub-saharan-africa-regional-edition.

Kamuzora, P. and L. Gilson (2007). Factors influencing implementation of the community health fund in Tanzania. *Health Policy and Planning 22*(2), 95–102.

Kenny, C. (2015). How the world bank should help the planet's poorest: Pay them. *Bloomberg*. http://www.bloomberg.com/news/articles/2015-04-21/how-the-world-bank-should-help-the-world-s-poorest-pay-them.

King, E. M. and J. R. Behrman (2009). Timing and duration of exposure in evaluations of social programs. *The World Bank Research Observer 24*(1), 55–82.

Leroy, J. L., M. Ruel, and E. Verhofstadt (2009). The impact of conditional cash transfer programmes on child nutrition: a review of evidence using a programme theory framework. *Journal of Development Effectiveness 1*(2), 103–129.

Macours, K., N. Schady, and R. Vakis (2008). Cash transfers, behavioral changes, and the cognitive development of young children: Evidence from a randomized experiment. *Policy Research Working Paper 4759*.

Manley, J., S. Gitter, and V. Slavchevska (2013). How effective are cash transfers at improving nutritional status? *World Development 48*, 133–155.

Marriott, A. (2011). Does health insurance work in Tanzania? http://www.globalhealthcheck.org/?p=388.

Morris, S. S., P. Olinto, R. Flores, E. A. Nilson, and A. C. Figueiro (2004). Conditional cash transfers are associated with a small reduction in the rate of weight gain of preschool children in northeast brazil. *The Journal of nutrition 134*(9), 2336–2341.

Özler, B. (2014). Obesity may not have dropped among children, but it almost certainly increased among the elderly. http://blogs.worldbank.org/impactevaluations/impact evaluations/impactevaluations/obesity-may-not-have-dropped-among-children-it-almost-certainly-increased-among-elderly.

Paxson, C. and N. Schady (2007). Cognitive development among young children in ecuador. *The Journal of Human Resources 42*(2), 49–84.

Paxson, C. and N. Schady (2010). Does money matter? the effects of cash transfers on child development in rural Ecuador. *Economic development and cultural change 59*(1), 187–229.

Powell-Jackson, T. and K. Hanson (2012). Financial incentives for maternal health: impact of a national programme in Nepal. *Journal of health economics 31*(1), 271–284.

Powell-Jackson, T., S. Mazumdar, and A. Mills (2015). Financial incentives in health: New evidence from India's Janani Suraksha Yojana. *Forthcoming, Journal of Health Economics*.

Robertson, L., P. Mushati, J. W. Eaton, L. Dumba, G. Mavise, J. Makoni, C. Schumacher, T. Crea, R. Monasch, L. Sherr, et al. (2013). Effects of unconditional and conditional cash transfers on child health and development in Zimbabwe: a cluster-randomised trial. *The Lancet 381*(9874), 1283–1292.

Schaffer, M. (2010). xtivreg2: Stata module to perform extended iv/2sls, gmm and ac/hac, liml and k-class regression for panel data models. http://ideas.repec.org/c/boc/bocode/s456501.html.

Strauss, J. and D. Thomas (1998). Health, nutrition, and economic development. *Journal of economic literature*, 766–817.

World Bank (2013). *Africa's Pulse 8*. http://www.worldbank.org/content/dam/Worldbank/document/Africa/Report/Africas-Pulse-brochure_Vol8.pdf.

World Health Organization (2014). United Republic of Tanzania: Health Profile. *Africa's Pulse*. http://www.who.int/gho/countries/tza.pdf?ua=1.

Table 1: Baseline (2009) means of outcomes and key characteristics by treatment assignment

	Treatment (T)		Control (C)		Difference (T-C)	
	Mean	N	Mean	N	Mean	S.E.
Panel A: Outcome variables						
Clinic Visits and Health, Full Sample						
# health clinic visits in last 12 months	2.83	3462	2.77	3456	0.06	(0.26)
Dummy - ill or injured in last four weeks	0.29	3462	0.26	3456	0.02	(0.02)
# of days unable to do normal activities	1.68	3462	1.59	3455	0.08	(0.14)
Clinic Visits and Health, 60 and over						
# health clinic visits in last 12 months	2.91	1049	2.67	1160	0.24	(0.35)
Dummy - ill or injured in last four weeks	0.39	1049	0.38	1160	0.01	(0.03)
# of days unable to do normal activities	2.78	1049	2.79	1159	-0.02	(0.31)
Clinic Visits and Health, 0-5 years old						
# health clinic visits in last 12 months	8.21	309	8.33	312	-0.12	(0.70)
Dummy - ill or injured in last four weeks	0.31	309	0.25	312	0.06*	(0.04)
# of days unable to do normal activities	1.31	309	0.80	312	0.51**	(0.21)
Health-related products						
Dummy - owns shoes	0.38	1515	0.47	1441	-0.10**	(0.05)
Dummy - owns slippers	0.62	1515	0.65	1441	-0.03	(0.03)
Insurance expenditures, thousands Tsh	0.26	881	0.11	879	0.15*	(0.08)
Specific activities of daily living, 60 and over						
Dummy - can do vigorous activity	0.35	1049	0.36	1160	-0.02	(0.04)
Dummy - can walk uphill	0.77	1049	0.75	1160	0.01	(0.03)
Dummy - can bend over or stoop	0.97	1049	0.96	1160	0.01	(0.01)
Dummy - can walk over 1km	0.86	1049	0.85	1160	0.01	(0.02)
Dummy - can walk over 100m	0.97	1049	0.96	1160	0.01	(0.01)
Dummy - can use bath or toilet	0.98	1049	0.97	1160	0.00	(0.01)
Ordinary activities index	4.89	1049	4.86	1160	0.03	(0.09)
Anthropometrics						
Height-for-age z-score	-1.46	231	-1.25	240	-0.21	(0.14)
Weight-for-age z-score	-0.90	208	-0.72	189	-0.18*	(0.10)
Weight-for-height z-score	0.06	187	0.04	176	0.02	(0.11)
BMI-for-age z-score	0.23	187	0.16	177	0.07	(0.12)
Height (cm)	87.38	234	87.10	241	0.28	(1.14)
Weight (kg)	12.22	253	12.07	253	0.14	(0.26)
MUAC (mm)	155.68	230	155.83	232	-0.15	(1.42)
Healthcare location						
None	0.17	993	0.13	907	0.04	(0.03)
Public: dispensary	0.38	993	0.41	907	-0.02	(0.04)
Public: district, region, or referal hospital	0.05	993	0.07	907	-0.02	(0.02)
Public: health center	0.03	993	0.04	907	0.00	(0.02)
Private: Pharmacy or chemist	0.28	993	0.27	907	0.01	(0.03)
Private: Healer, herbalist, or faith healer	0.05	993	0.05	907	0.00	(0.01)
Private: dispensary, hospital, clinic, or store	0.02	993	0.02	907	0.00	(0.01)
Private: Mission dispensary or hospital	0.01	993	0.01	907	0.00	(0.01)
Health Facility Rating						
Dummy - share rating facility as good or excellent	0.70	880	0.72	879	-0.02	(0.03)
Dummy - share rating facility as excellent	0.05	880	0.06	879	-0.01	(0.02)
Household size						
Total	3.94	879	3.94	878	0.00	(0.21)
0-5 years	0.35	879	0.36	878	0.00	(0.04)
60 and over	1.19	879	1.32	878	-0.13**	(0.05)
Healthcare financing method						
Free treatment	0.20	885	0.23	826	-0.03	(0.03)
Loan or assistance	0.11	885	0.10	826	0.01	(0.02)
Cash or asset	0.65	885	0.66	826	0.00	(0.03)
Health insurance	0.04	885	0.02	826	0.02	(0.02)

table continued on next page...

	Treatment (T)		Control (C)		Difference (T-C)	
	Mean	N	Mean	N	Mean	S.E.
Panel B: Individual and household characteristics						
Individual characteristics						
Age	35.54	3462	37.04	3456	-1.49	(1.20)
Dummy - male	0.47	3462	0.45	3456	0.02	(0.01)
Dummy - has less than Standard 1 education	0.53	3459	0.54	3451	0.00	(0.02)
Dummy - has Standard 1-4 education	0.22	3459	0.22	3451	0.00	(0.01)
Dummy - has at least Standard 5 education	0.24	3459	0.24	3451	0.00	(0.02)
Dummy - literate	0.41	3462	0.42	3456	-0.01	(0.03)
Household characteristics						
Dummy - household has improved roof	0.33	880	0.37	878	-0.04	(0.06)
Dummy - household has improved floor	0.03	880	0.09	878	-0.06**	(0.02)
Dummy - household has toilet facilities	0.69	880	0.76	879	-0.07	(0.04)
Dummy - household has piped water	0.30	880	0.32	879	-0.01	(0.08)
Dummy - head of household is male	0.63	879	0.59	878	0.04	(0.03)

Source: Authors' calculations based on baseline (2009) household survey data.

Notes: Treatment indicates assignment to treatment. The universe of individuals used to summarize the healthcare location and healthcare financing method outcomes is individuals who reported being ill or injured in the last four weeks. Ordinary activities index is the sum of the six activity dummies; its range is 0 to 6. Specific activities of daily living are summarized for those at least 60 years old because data were unavailable for individuals under 60 years old at the time of the midline and endline surveys. Shoe and slipper ownership are summarized for those 18 years old or younger in the baseline survey. Insurance expenditures is a household level outcome, and it refers to total annual medical, car, and life insurance expenditures. BMI is body mass index and MUAC is middle upper-arm circumference. Ages refer to age at time of baseline survey. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table 2: Effects of treatment on health clinic visits of the treated in the last 12 months

	Full sample: baseline and endline (1)	0-5 years old (2)	60 and over (3)
Panel A: Effect of treatment on the treated			
Treatment × 2011 (midline)		2.637***	1.138***
		(0.966)	(0.361)
Treatment × 2012 (endline)	-0.072	-1.153	0.170
	(0.268)	(0.954)	(0.359)
2011 (midline)		-3.812***	-1.217***
		(0.580)	(0.215)
2012 (endline)	-1.436***	-5.751***	-0.671***
	(0.181)	(0.632)	(0.236)
R^2	0.061	0.374	0.018
Baseline mean	2.802	8.272	2.783
Observations	9622	1092	5339
Panel B: Heterogeneous treatment effects by staff/capita			
Treatment effect for those with fewer (midline)		3.388***	0.908*
		(1.126)	(0.488)
Treatment effect for those with more (midline)		1.650	1.366**
		(1.706)	(0.536)
Treatment effect for those with fewer (endline)	-0.001	-1.637	0.207
	(0.389)	(1.212)	(0.527)
Treatment effect for those with more (endline)	-0.157	-0.259	0.135
	(0.359)	(1.348)	(0.493)
p-value of difference (midline)		0.395	0.527
p-value of difference (endline)	0.768	0.447	0.920
Baseline mean for those with fewer	2.598	7.553	2.501
Baseline mean for those with more	3.004	9.104	3.051

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.

Notes: Midline data are excluded from the full sample because health facility visit data were not collected in the midline survey for those 5-60 years old. Ages refer to age at the time of baseline survey. Fewer refers to those residing in villages in the bottom half of the distribution of baseline health clinic staff per capita, while more refers to those in the top half. All specifications include individual fixed effects. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table 3: Effects of treatment on take-up of health-related products

	Dummy - owns shoes	Dummy - owns slippers	Insurance expenditures (thousands Tsh)
	(1)	(2)	(3)
Panel A: Effect of treatment on the treated			
Treatment × 2011 (midline)	0.194***	0.058	1.262***
	(0.047)	(0.037)	(0.267)
Treatment × 2012 (endline)	0.192***	0.089**	1.623***
	(0.050)	(0.041)	(0.302)
2011 (midline)	0.128***	0.188***	0.172***
	(0.028)	(0.023)	(0.051)
2012 (endline)	0.125***	0.195***	0.433***
	(0.031)	(0.028)	(0.099)
R^2	0.106	0.106	0.124
Baseline mean	0.423	0.632	0.181
Observations	6138	6138	4920
Panel B: Heterogeneous treatment effects by degree of poverty			
Treatment effect for extremely poor (midline)	0.318***	0.101*	1.469***
	(0.069)	(0.054)	(0.302)
Treatment effect for moderately poor (midline)	0.086*	0.016	0.962***
	(0.048)	(0.038)	(0.281)
Treatment effect for extremely poor (endline)	0.262***	0.170***	1.842***
	(0.070)	(0.052)	(0.319)
Treatment effect for moderately poor (endline)	0.129**	0.009	1.337***
	(0.058)	(0.049)	(0.412)
p-value of difference (midline)	0.002	0.137	0.084
p-value of difference (endline)	0.099	0.009	0.226
Baseline mean for extremely poor	0.292	0.571	0.093
Baseline mean for moderately poor	0.525	0.680	0.289

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.

Notes: Shoe and slipper ownership are individual-level outcomes for those at least 18 years old at the time of the baseline survey. Insurance expenditures is a household level outcome, and it refers to total annual medical, car, and life insurance expenditures. Degree of poverty refers to the value at the time of the baseline survey on an index of asset ownership. The index is the first principal component from a PCA using information on ownership of 13 household assets. Extremely poor refers to those in the bottom half, while moderately poor refers to those in the top half. Columns (1) and (2) include individual fixed effects. Column (3) includes household fixed effects. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table 4: Effects of treatment on illness and injury in the last four weeks

	Dummy - ill or injured in last four weeks			Days in last four weeks unable to perform normal daily activities due to illness or injury		
	Full sample (1)	0-5 years old (2)	60 and over (3)	Full sample (4)	0-5 years old (5)	60 and over (6)
Panel A: Effect of treatment on the treated						
Treatment × 2011 (midline)	0.004	-0.012	0.046	-0.225	-0.135	-0.215
	(0.027)	(0.060)	(0.042)	(0.240)	(0.311)	(0.511)
Treatment × 2012 (endline)	-0.046*	-0.118*	-0.002	-0.465**	-0.835**	-0.371
	(0.027)	(0.068)	(0.036)	(0.234)	(0.390)	(0.433)
2011 (midline)	0.002	-0.054*	0.032	0.199	-0.205	0.675**
	(0.018)	(0.032)	(0.028)	(0.165)	(0.171)	(0.322)
2012 (endline)	0.078***	0.032	0.147***	1.077***	0.306	2.390***
	(0.015)	(0.047)	(0.023)	(0.147)	(0.297)	(0.268)
R^2	0.007	0.013	0.024	0.012	0.010	0.033
Baseline mean	0.276	0.282	0.388	1.636	1.052	2.786
Observations	18180	1431	5341	18180	1431	5341
Panel B: Heterogeneous treatment effects by staff/capita						
Treatment effect for those with fewer (midline)	0.016	-0.012	0.014	0.087	-0.129	0.227
	(0.040)	(0.069)	(0.062)	(0.331)	(0.355)	(0.771)
Treatment effect for those with more (midline)	-0.007	-0.023	0.077	-0.539	-0.204	-0.619
	(0.035)	(0.094)	(0.058)	(0.337)	(0.494)	(0.659)
Treatment effect for those with fewer (endline)	-0.030	-0.120	0.009	0.075	-0.532	0.800
	(0.041)	(0.090)	(0.049)	(0.321)	(0.418)	(0.543)
Treatment effect for those with more (endline)	-0.061*	-0.124	-0.020	-1.045***	-1.262*	-1.604***
	(0.033)	(0.102)	(0.051)	(0.306)	(0.721)	(0.590)
p-value of difference (midline)	0.667	0.928	0.458	0.185	0.901	0.404
p-value of difference (endline)	0.553	0.979	0.683	0.012	0.381	0.003
Baseline mean for those with fewer	0.284	0.297	0.380	1.662	1.081	2.763
Baseline mean for those with more	0.267	0.264	0.396	1.610	1.017	2.809

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.
Notes: Ages refer to age at the time of baseline survey. Fewer refers to those residing in villages in the bottom half of the distribution of baseline health clinic staff per capita, while more refers to those in the top half. All specifications include individual fixed effects. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table 5: Effects of treatment on ability to perform ordinary activities

	dummy - can ...						
	do vigorous activity	walk uphill	bend over or stoop	walk more than 1km	walk more than 100m	use bath or toilet	Ordinary activities index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment × 2011 (midline)	0.027	-0.012	-0.004	-0.007	-0.018*	0.003	0.003
	(0.051)	(0.041)	(0.009)	(0.024)	(0.011)	(0.007)	(0.106)
Treatment × 2012 (endline)	0.013	-0.023	0.013	-0.027	-0.016**	0.013	-0.050
	(0.062)	(0.037)	(0.014)	(0.024)	(0.007)	(0.012)	(0.104)
2011 (midline)	0.188***	0.078***	-0.005	0.028*	0.003	-0.001	0.289***
	(0.035)	(0.025)	(0.006)	(0.015)	(0.008)	(0.004)	(0.065)
2012 (endline)	-0.177***	-0.027	-0.046***	-0.049***	0.011**	-0.027***	-0.050
	(0.042)	(0.026)	(0.010)	(0.018)	(0.005)	(0.009)	(0.070)
R^2	0.156	0.027	0.018	0.025	0.003	0.010	0.059
Baseline mean	0.356	0.760	0.968	0.855	0.962	0.974	4.875
Observations	5329	5329	5329	5329	5040	5329	5040

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.
Notes: Activity index is the sum of the six activity dummies; its range is 0 to 6. Only those at least 60 years old at the time of the baseline are included, due to data availability. All specifications include individual fixed effects. Treatment estimates are estimates of the effect of treatment on the treated. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table 6: Effects of treatment on anthropometrics for children aged 0-5

	z-scores						
	Height-for-age z-score	Weight-for-age z-score	Weight-for-height z-score	BMI-for-age z-score	Height (cm)	Weight (kg)	MUAC (mm)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment × 2011 (midline)	0.115	-0.025	-0.323	-0.405	-0.806	-0.204	-1.602
	(0.268)	(0.238)	(0.343)	(0.357)	(1.306)	(0.253)	(2.604)
Treatment × 2012 (endline)	0.253	-0.128	-0.481	-0.551	-0.746	0.012	0.192
	(0.409)	(0.241)	(0.365)	(0.408)	(1.443)	(0.300)	(2.971)
2011 (midline)	0.000	0.228*	0.504***	0.500**	-0.117	0.554***	-0.619
	(0.180)	(0.138)	(0.189)	(0.209)	(0.481)	(0.155)	(1.471)
2012 (endline)	0.607**	0.806***	0.559***	0.483**	0.271	0.718***	0.099
	(0.262)	(0.141)	(0.208)	(0.241)	(0.829)	(0.180)	(1.615)
R^2	0.081	0.087	0.078	0.076	0.064	0.074	0.039
Baseline mean	-1.354	-0.812	0.052	0.197	87.240	12.146	155.753
Observations	967	983	848	841	1008	1183	998

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.
Notes: Regressions include village × cohort fixed effects rather than individual fixed effects. Cohorts included are the following, defined in terms of current age at the time of each survey round: 0-6 months, 7-12 months, 13-18 months, 19-24 months, 25-30 months, 31-36 months, 37-42 months, 43-48 months, 49-54 months, and 55-60 months. Baseline controls not shown include the age, age^2, sex, and education level of the household head. Also included are dummies for gender, household size, having an improved roof, having an improved toilet, having an improved floor, having piped water, village population, and the first principal components from a PCA using information on ownership of 13 household assets at baseline. BMI is body mass index and MUAC is middle upper-arm circumference. Children with z-scores less than -6.0 or greater than 6.0 were excluded from the analysis; 59 of 1,246 height-for-age z-scores were excluded; 53 of 1,260 weight-for-age z-scores were excluded; 11 of 1,093 weight-for-height z-scores were excluded; and 14 of 1,090 BMI-for-age z-scores were excluded. Treatment estimates are estimates of the effect of treatment on the treated. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table 7: Effects of treatment on participation in the government-run health insurance program, the community health fund (CHF)

	Dummy - participates in the CHF (1)	Dummy - reports an inability to afford the CHF (2)
Panel A: Effect of treatment on the treated		
Treatment × 2012 (endline)	0.379***	-0.434***
	(0.040)	(0.039)
R^2	0.313	0.230
Observations	1555	1064
Panel B: Heterogeneous treatment effects by degree of poverty		
Treatment effect for extremely poor (endline)	0.398***	-0.454***
	(0.049)	(0.045)
Treatment effect for moderately poor (endline)	0.355***	-0.410***
	(0.039)	(0.043)
p-value of difference (endline)	0.302	0.297

Source: Authors' calculations based on endline (2012) household survey data.
Notes: Data on these outcomes are only available from the endline survey, and are at the household level. The universe for column 1 is all households. Households that report having never heard of the CHF are assumed to not be participating in the CHF. The universe for column 2 is households that report ever hearing of the CHF. Households that particiapte in the CHF are considered able to afford the CHF. Baseline controls not shown include the age, age^2, sex, and education level of the household head. Also included are dummies for district, household size, having an improved roof, having an improved toilet, having an improved floor, having piped water, village population, the number of years since the CHF began operating in respondent's village, and the first principal components from a PCA using information on ownership of 13 household assets at baseline. Degree of poverty refers to the value at the time of the baseline survey on an index of asset ownership. The index is the first principal component from a PCA using information on ownership of 13 household assets. Extremely poor refers to those in the bottom half, while moderately poor refers to those in the top half. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table 8: Effects of treatment on type of health provider visited to address main health problem of the last four weeks

		Public			Private			
	None	Dispensary	District, region, or referal hospital	Health center	Pharmacy or chemist	Healer, herbalist, or faith healer	Dispensary, hospital, clinic, or store	Mission dispensary or hospital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Effect of treatment on the treated								
Treatment × 2011 (midline)	-0.127***	0.186***	-0.039	0.001	-0.037	0.026	-0.008	-0.002
	(0.043)	(0.050)	(0.025)	(0.027)	(0.043)	(0.017)	(0.018)	(0.008)
Treatment × 2012 (endline)	-0.058	0.158***	-0.004	-0.023	-0.059	-0.010	-0.001	-0.004
	(0.039)	(0.052)	(0.030)	(0.027)	(0.050)	(0.024)	(0.020)	(0.015)
2011 (midline)	0.101***	-0.033	0.010	0.030	-0.066**	-0.027*	-0.011	-0.003
	(0.032)	(0.035)	(0.019)	(0.020)	(0.033)	(0.014)	(0.012)	(0.004)
2012 (endline)	0.005	0.052	0.017	0.029	-0.088**	-0.012	-0.004	0.001
	(0.027)	(0.041)	(0.025)	(0.020)	(0.041)	(0.017)	(0.015)	(0.013)
R^2	0.015	0.032	0.004	0.006	0.024	0.002	0.002	0.001
Baseline mean	0.154	0.395	0.059	0.035	0.274	0.047	0.023	0.013
Observations	2929	2929	2929	2929	2929	2929	2929	2929
Panel B: Heterogeneous treatment effects by degree of poverty								
Treatment effect for extremely poor (midline)	-0.105*	0.222***	-0.048	-0.056	0.027	0.001	-0.046**	0.005
	(0.060)	(0.075)	(0.033)	(0.035)	(0.069)	(0.025)	(0.023)	(0.014)
Treatment effect for moderately poor (midline)	-0.149***	0.149*	-0.032	0.066*	-0.114*	0.055**	0.035	-0.011
	(0.055)	(0.080)	(0.038)	(0.039)	(0.064)	(0.028)	(0.028)	(0.016)
Treatment effect for extremely poor (endline)	-0.017	0.171***	-0.044	-0.032	-0.025	-0.023	-0.035	0.006
	(0.052)	(0.063)	(0.039)	(0.033)	(0.065)	(0.032)	(0.022)	(0.022)
Treatment effect for moderately poor (endline)	-0.101*	0.142*	0.037	-0.006	-0.100	0.006	0.038	-0.015
	(0.056)	(0.076)	(0.040)	(0.036)	(0.075)	(0.038)	(0.033)	(0.016)
p-value of difference (midline)	0.553	0.533	0.751	0.017	0.169	0.158	0.016	0.529
p-value of difference (endline)	0.260	0.749	0.117	0.541	0.456	0.564	0.057	0.389
Baseline mean for extremely poor	0.190	0.381	0.029	0.044	0.268	0.060	0.016	0.012
Baseline mean for moderately poor	0.120	0.409	0.087	0.028	0.279	0.036	0.029	0.013

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.
Notes: Over the 3 rounds of the survey, respondents reported being sick or injured a total of 5,922 times. In all of those reports, the most important health provider was reported. 33 people were excluded from this analysis for reporting that the most important health provider was other. Individuals who were only sick once are not counted in the number of observations. Degree of poverty refers to the value at the time of the baseline survey on an index of asset ownership. The index is the first principal component from a PCA using information on ownership of 13 household assets. Extremely poor refers to those in the bottom half, while moderately poor refers to those in the top half. All specifications include individual fixed effects. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table 9: Effects of treatment on reported household satisfaction with community health facilities

	Dummy - share rating facility as...	
	good or excellent	excellent
	(1)	(2)
Treatment × 2011 (midline)	0.137***	0.044**
	(0.051)	(0.018)
Treatment × 2012 (endline)	0.031	0.039
	(0.043)	(0.025)
2011 (midline)	-0.132***	-0.045***
	(0.035)	(0.013)
2012 (endline)	-0.120***	-0.002
	(0.026)	(0.017)
R^2	0.018	0.010
Baseline mean	0.709	0.055
Observations	4920	4920

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.
Notes: In all three survey rounds, households were asked to rate the health facilities in their community as poor, fair, good, or excellent. All specifications include household fixed effects. Treatment estimates are estimates of the effect of treatment on the treated. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table 10: Effects of treatment on objective health facility quality measures

	Infrastructure quality index	Staffing and treatment quality index	Overall health facility quality index
	(1)	(2)	(3)
Treatment village	-0.117	-0.315	-0.306
	(0.260)	(0.238)	(0.297)
Number of doctors, 2009	0.355***	0.460***	0.581***
	(0.089)	(0.130)	(0.137)
Number of nurses, 2009	0.049	0.037	0.064
	(0.039)	(0.065)	(0.069)
Number of other assistants/staff, 2009	0.043***	0.054***	0.070***
	(0.007)	(0.010)	(0.012)
ln(population, 2009)	0.016	0.074	0.053
	(0.166)	(0.251)	(0.263)
Constant	-1.173	-1.552	-1.839
	(1.361)	(1.945)	(2.080)
Observations	80	80	80

Source: Authors' calculations based on endline (2012) household survey data.
Notes: District fixed effects are included but not shown. Each index is the first principal component from a PCA including all of the measures capturing the stated dimension of quality (infrastructure, staffing and treatment, or overall). The infrastructure quality measures include the number of rooms to treat patients in the facility, the number of other rooms in the facility, the number of impatient beds in the facility, a dummy for whether there is a Community Health Fund operating in the village, and dummies for whether the facility has: a laboratory, bed for maternal deliveries, operating room, piped water, motorized vehicle, an incinerator, refrigerator, and electricity. The staffing and treatment quality measures include the number of patients referred to another facility, the number of patients not able to be treated as much as desired, the number of patients seen in the past 30 days, the number of hours in the last week that a doctor was attending to patients, the number of doctors, the number of nurses, the number of assistants/staff, and dummies for whether the facility: provides HIV services beyond testing, has a compulsory vaccination program, has family planning services, has a childhood nutrition program, tests for HIV, tests for typhoid, tests for malaria, tests for anemia, tests for worms, tests blood sugar, tests blood pressure, and tests for typhoid. The overall health facility quality index is constructed using the union of these two groups of measures. If a village has more than one facility, then data for the primary facility are used. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Appendix Tables

Table A1: Effects of treatment on mortality

	Full sample (1)	0-5 years old (2)	60 and over (3)
Treatment	0.006	-0.004	0.011
	(0.003)	(0.006)	(0.009)
2012 (endline)	-0.025***	-0.010**	-0.066***
	(0.003)	(0.005)	(0.007)
R^2	0.065	0.032	0.065
Observations	13042	1192	3993

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.

Notes: The analysis includes (up to) two observations per individual: a midline observation (which examines deaths between baseline and midline) and an endline observation (which examines deaths between midline and endline). At midline, the outcome is a dummy for the individual being dead at midline (individuals not in sample at baseline take on a missing value). At endline, the outcome is a dummy for the individual being dead at endline (individuals not in sample at midline take on a missing value). Baseline controls not shown, include the age, age^2, sex, and education level of both the respondent and the household head. Also included are district fixed effects and dummies for gender, household size, having an improved roof, having an improved toilet, having an improved floor, having piped water, village population, and the first principal components from a PCA using information on ownership of 13 household assets at baseline. Ages refer to age at the time of baseline survey. Treatment estimates are estimates of the effect of treatment on the treated. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table A2: Effects of treatment on household size

	Full sample (1)	0-5 years old (2)	60 and over (3)
Treatment × 2011 (midline)	0.041	0.022	0.030
	(0.110)	(0.038)	(0.031)
Treatment × 2012 (endline)	-0.018	0.018	0.014
	(0.131)	(0.046)	(0.037)
2011 (midline)	0.034	-0.073***	-0.074***
	(0.079)	(0.024)	(0.021)
2012 (endline)	0.134	-0.065**	-0.040
	(0.104)	(0.031)	(0.025)
R^2	0.002	0.005	0.006
Baseline mean	3.937	0.353	1.257
Observations	4913	4913	4913

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.

Notes: Ages refer to age at the time of baseline survey. All specifications include household fixed effects. Treatment estimates are estimates of the effect of treatment on the treated. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table A3: Robustness of results to Poisson estimation for count data models

estimate	outcome	column	OLS: S.E. clustered at village level, estimate (1)	OLS: S.E. clustered at village level, S.E. (2)	Poisson: S.E. clustered at village level, estimate (3)	Poisson: S.E. clustered at village level, S.E. (4)	Poisson: bootstrap S.E., estimate (5)	Poisson: bootstrap S.E., S.E. (6)
Panel A: Effects of treatment on health clinic visits of the treated in the last 12 months (table 2)								
Treatment × 2011	Total	1	.	(.)	.	(.)	.	(.)
Treatment × 2012	Total	1	-0.072	(0.268)	0.003	(0.104)	0.003	(0.103)
Treatment × 2011	0-5 years	2	2.637***	(0.966)	0.385***	(0.137)	0.385*	(0.224)
Treatment × 2012	0-5 years	2	-1.153	(0.954)	-0.392	(0.273)	-0.392	(0.299)
Treatment × 2011	60 and over	3	1.138***	(0.362)	0.578***	(0.138)	0.578***	(0.122)
Treatment × 2012	60 and over	3	0.170	(0.359)	0.098	(0.134)	0.098	(0.133)
Panel B: Effects of treatment on number of days unable to perform normal daily activities (table 4)								
Treatment × 2011	Total	1	-0.225	(0.240)	-0.140	(0.149)	-0.140	(0.122)
Treatment × 2012	Total	1	-0.465**	(0.234)	-0.236**	(0.112)	-0.236*	(0.122)
Treatment × 2011	0-5 years	2	-0.135	(0.311)	-0.008	(0.345)	-0.008	(0.431)
Treatment × 2012	0-5 years	2	-0.835**	(0.390)	-0.832**	(0.387)	-0.832*	(0.450)
Treatment × 2011	60 and over	3	-0.215	(0.511)	-0.091	(0.175)	-0.091	(0.163)
Treatment × 2012	60 and over	3	-0.371	(0.433)	-0.119	(0.125)	-0.119	(0.158)
Panel C: Effects of treatment on household size (table A2)								
Treatment × 2011	Total	1	0.041	(0.110)	0.010	(0.027)	0.010	(0.031)
Treatment × 2012	Total	1	-0.018	(0.131)	-0.004	(0.031)	-0.004	(0.034)
Treatment × 2011	0-5 years	2	0.022	(0.038)	0.064	(0.113)	0.064	(0.126)
Treatment × 2012	0-5 years	2	0.018	(0.046)	0.053	(0.135)	0.053	(0.142)
Treatment × 2011	60 and over	3	0.030	(0.031)	0.019	(0.024)	0.019	(0.029)
Treatment × 2012	60 and over	3	0.014	(0.037)	0.008	(0.029)	0.008	(0.031)

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.
Notes: Standard errors in column 6 are bootstrapped over 1000 samples with replacement. The first row of Panel A has missing data since health clinic visit data for those aged 5 - 60 were not collected at midline. Column refers to the column in which the estimate appears in the original table. All specifications include individual fixed effects. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table A4: Effects of treatment on method used to finance healthcare when addressing main health problem of the last four weeks

	Free treatment (1)	Loan or assistance (2)	Cash or asset (3)	Health insurance (4)
Panel A: Effect of treatment on the treated				
Treatment × 2011 (midline)	0.039	-0.018	-0.190***	0.169***
	(0.055)	(0.036)	(0.066)	(0.050)
Treatment × 2012 (endline)	-0.089*	-0.015	-0.197***	0.300***
	(0.052)	(0.034)	(0.070)	(0.057)
2011 (midline)	-0.007	-0.071***	0.054	0.023
	(0.035)	(0.027)	(0.037)	(0.015)
2012 (endline)	0.027	-0.074***	0.007	0.040***
	(0.037)	(0.023)	(0.039)	(0.013)
R^2	0.006	0.035	0.025	0.142
Baseline mean	0.216	0.103	0.654	0.027
Observations	2478	2478	2478	2478
Panel B: Heterogeneous treatment effects by degree of poverty				
Treatment effect for extremely poor (midline)	0.059	-0.073	-0.153	0.166**
	(0.084)	(0.048)	(0.101)	(0.070)
Treatment effect for moderately poor (midline)	0.022	0.036	-0.230***	0.171***
	(0.058)	(0.050)	(0.077)	(0.053)
Treatment effect for extremely poor (endline)	-0.121	-0.074*	-0.153	0.348***
	(0.075)	(0.042)	(0.098)	(0.073)
Treatment effect for moderately poor (endline)	-0.060	0.047	-0.242***	0.255***
	(0.064)	(0.050)	(0.079)	(0.055)
p-value of difference (midline)	0.692	0.090	0.530	0.940
p-value of difference (endline)	0.517	0.049	0.430	0.196
Baseline mean for extremely poor	0.247	0.128	0.610	0.015
Baseline mean for moderately poor	0.188	0.081	0.693	0.038

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.

Notes: Over the 3 rounds of the survey, respondents reported being sick or injured a total of 5,922 times. In 5,409 of those 5,922 reports, the main treatment financing method was reported. 44 people were excluded from this analysis for reporting financing with either 'other' or differed by provider since it was not possible to understand how such individuals financed treatment. Individuals who were only sick once are not counted in the number of observations. Degree of poverty refers to the value at the time of the baseline survey on an index of asset ownership. The index is the first principal component from a PCA using information on ownership of 13 household assets. Extremely poor refers to those in the bottom half, while moderately poor refers to those in the top half. All specifications include individual fixed effects. Standard errors are in parentheses and clustered at the village level. *** indicates p<0.01; ** indicates p<0.05; and * indicates p<0.10.

Table A5: Robustness testing: multiple hypothesis testing

treatment estimate	outcome	column	estimate	p-value	BKY***	BH**	Bon.*
Panel A: Effects of treatment on health clinic visits of the treated in the last 12 months (table 2)							
Treatment (T) × 2011	0-5 years	2	2.64	0.01	0.01	0.01	0.04
T × 2011	60 and over	3	1.14	0.00	0.01	0.01	0.01
T × 2011 × fewer staff/capita	0-5 years	2	3.39	0.00	0.01	0.01	0.02
T × 2011 × fewer staff/capita	60 and over	3	0.91	0.06	0.03	0.08	0.38
T × 2011 × more staff/capita	60 and over	3	1.37	0.01	0.01	0.02	0.06
Panel B: Effects of treatment on uptake of health-related products (table 3)							
T × 2011	Dummy - owns shoes	1	0.19	0.00	0.00	0.00	0.00
T × 2012	Dummy - owns shoes	1	0.19	0.00	0.00	0.00	0.00
T × 2012	Dummy - owns slippers	2	0.09	0.03	0.02	0.03	0.17
T × 2011	Insurance expenditures, thousands	3	1.26	0.00	0.00	0.00	0.00
T × 2012	Insurance expenditures, thousands	3	1.62	0.00	0.00	0.00	0.00
T × 2011 × extremely poor	Dummy - owns shoes	1	0.32	0.00	0.00	0.00	0.00
T × 2011 × moderately poor	Dummy - owns shoes	1	0.09	0.07	0.08	0.11	0.44
T × 2012 × extremely poor	Dummy - owns shoes	1	0.26	0.00	0.00	0.00	0.00
T × 2012 × moderately poor	Dummy - owns shoes	1	0.13	0.03	0.02	0.03	0.16
T × 2011 × extremely poor	Dummy - owns slippers	2	0.10	0.06	0.08	0.11	0.39
T × 2012 × extremely poor	Dummy - owns slippers	2	0.17	0.00	0.00	0.00	0.01
T × 2011 × extremely poor	Insurance expenditures, thousands	3	1.47	0.00	0.00	0.00	0.00
T × 2011 × moderately poor	Insurance expenditures, thousands	3	0.96	0.00	0.00	0.00	0.00
T × 2012 × extremely poor	Insurance expenditures, thousands	3	1.84	0.00	0.00	0.00	0.00
T × 2012 × moderately poor	Insurance expenditures, thousands	3	1.34	0.00	0.00	0.00	0.00
Panel C: Effects of treatment on illness and injury in the last four weeks (table 4)							
T × 2012	Dummy - ill or injured in last four weeks	1	-0.05	0.10	0.40	0.29	0.86
T × 2012	Dummy - ill or injured in last four weeks	2	-0.12	0.08	0.40	0.29	0.75
T × 2012	# of days unable to do normal activities	4	-0.46	0.05	0.09	0.11	0.43
T × 2012	# of days unable to do normal activities	5	-0.83	0.03	0.08	0.10	0.29
T × 2012 × more staff/capita	Dummy - ill or injured in last four weeks	1	-0.06	0.07	0.40	0.29	0.59
T × 2012 × more staff/capita	# of days unable to do normal activities	4	-1.04	0.00	0.01	0.01	0.01
T × 2012 × more staff/capita	# of days unable to do normal activities	5	-1.26	0.08	0.11	0.14	0.72
T × 2012 × more staff/capita	# of days unable to do normal activities	6	-1.60	0.01	0.03	0.03	0.06
Panel D: Effects of treatment on ability to perform ordinary activities (table 5)							
T × 2011	walk more than 100m	5	-0.02	0.10	1.00	0.59	0.59
T × 2012	walk more than 100m	5	-0.02	0.02	0.12	0.11	0.11
Panel E: Effects of treatment on participation in the government-run health insurance program, the community health fund (CHF) (table 7)							
T × 2012	Dummy - participates in the CHF	1	0.38	0.00	0.00	0.00	0.00
T × 2012	Dummy - unable to afford the CHF	2	-0.43	0.00	0.00	0.00	0.00
T × 2012 × moderately poor	Dummy - participates in the CHF	1	0.36	0.00	0.00	0.00	0.00
T × 2012 × extremely poor	Dummy - participates in the CHF	1	0.40	0.00	0.00	0.00	0.00
T × 2012 × moderately poor	Dummy - unable to afford the CHF	2	-0.41	0.00	0.00	0.00	0.00
T × 2012 × extremely poor	Dummy - unable to afford the CHF	2	-0.45	0.00	0.00	0.00	0.00

table continued on next page...

treatment estimate	outcome	column	estimate	p-value	BKY***	BH**	Bon.*
Panel F: Effects of treatment on type of health provider visited to address main health problem of the last four weeks (table 8)							
T × 2011	None	1	-0.13	0.00	0.03	0.03	0.08
T × 2011	Public: Dispensary	2	0.19	0.00	0.01	0.01	0.00
T × 2012	Public: Dispensary	2	0.16	0.00	0.06	0.06	0.06
T × 2011 × moderately poor	None	1	-0.15	0.01	0.04	0.04	0.16
T × 2011 × extremely poor	None	1	-0.10	0.08	0.21	0.21	1.91
T × 2012 × moderately poor	None	1	-0.10	0.07	0.64	0.43	1.70
T × 2011 × extremely poor	Public: Dispensary	2	0.22	0.00	0.03	0.03	0.07
T × 2011 × moderately poor	Public: Dispensary	2	0.15	0.06	0.21	0.21	1.52
T × 2012 × extremely poor	Public: Dispensary	2	0.17	0.01	0.08	0.08	0.16
T × 2012 × moderately poor	Public: Dispensary	2	0.14	0.06	0.64	0.43	1.48
T × 2011 × moderately poor	Public: Health center	4	0.07	0.09	0.21	0.21	2.09
T × 2011 × moderately poor	Private: Pharmacy or chemist	5	-0.11	0.08	0.21	0.21	1.86
T × 2011 × moderately poor	Private: Healer, herbalist, or faith healer	6	0.06	0.05	0.18	0.19	1.11
T × 2011 × extremely poor	Private: Dispensary, hospital, clinic, or store	7	-0.05	0.04	0.18	0.19	1.01
Panel G: Effects of treatment on reported household satisfaction with community health facilities (table 9)							
T × 2011	Dummy - share rating facility as good or excellent	1	0.14	0.01	0.01	0.01	0.02
T × 2011	Dummy - share rating facility as excellent	2	0.04	0.01	0.01	0.01	0.03
Panel H: Effects of treatment on method used to finance healthcare when addressing main health problem of the last four weeks (table A4)							
T × 2012	Free treatment	1	-0.09	0.09	0.10	0.15	1.07
T × 2011	Cash or asset	3	-0.19	0.00	0.01	0.01	0.05
T × 2012	Cash or asset	3	-0.20	0.00	0.01	0.01	0.06
T × 2011	Health insurance	4	0.17	0.00	0.01	0.01	0.01
T × 2012	Health insurance	4	0.30	0.00	0.00	0.00	0.00
T × 2012 × extremely poor	Loan or assistance	2	-0.07	0.08	0.10	0.15	1.00
T × 2011 × moderately poor	Cash or asset	3	-0.23	0.00	0.01	0.01	0.03
T × 2012 × moderately poor	Cash or asset	3	-0.24	0.00	0.00	0.01	0.02
T × 2011 × extremely poor	Health insurance	4	0.17	0.02	0.03	0.04	0.21
T × 2011 × moderately poor	Health insurance	4	0.17	0.00	0.01	0.01	0.01
T × 2012 × moderately poor	Health insurance	4	0.26	0.00	0.00	0.00	0.00
T × 2012 × extremely poor	Health insurance	4	0.35	0.00	0.00	0.00	0.00

Source: Authors' calculations based on baseline (2009), midline (2011), and endline (2012) household survey data.
Notes: Treatment effect estimates with p-values<0.10 displayed. Midline and endline treatment effects are abbreviated T × 2011 and T × 2012, respectively. Column refers to the column in which the estimate appears in the original table. ***BKY stands for Benjamini, Krieger, and Yekutieli q-values. **BH stands for Benjamini and Hochberg q-values. Q-values represent the smallest level at which the hypothesis is rejected. *Bon. stands for Bonerferroni p-values (p-value × number of hypotheses). BKY, BH, and Bon. are estimated separately for each period and table. The three exceptions to this rule are table 3, table 4, and table 5. In table 3, the two individual-level outcomes (owning shoes and slippers) are grouped separately from insurance expenditures (household-level). In table 4, the two outcomes were grouped separately (one group for columns 1-3 and another group for columns 4-6; one measures the extensive margin of illness, while the other measures the intensive margin). In table 3, the 6 specific activities of daily living are considered separately from the ordinary activities index; estimation of an index serves as an additional check on our corrections for multiple inference. Tables 6, 10, A2, and A1 do not appear in this table because there were no treatment estimates with a p-value less than 0.10 in those tables.