

Shoeing the Children

The Impact of the TOMS Shoe Donation Program in Rural El Salvador

Bruce Wydick

Elizabeth Katz

Flor Calvo

Felipe Gutierrez

Brendan Janet



WORLD BANK GROUP

Development Economics Vice Presidency

Operations and Strategy Team

September 2016

Abstract

The study uses a cluster-randomized trial among 1,578 children from 979 households in rural El Salvador to test the impacts of TOMS shoe donations on children's time allocation, school attendance, health, self-esteem, and aid dependency. Results indicate high levels of usage and approval of the shoes by children in the treatment group, and time diaries show modest evidence that the donated shoes allocated children's time toward outdoor activities. Difference-in-difference and ANCOVA estimates

find generally insignificant impacts on overall health, foot health, and self-esteem but small positive impacts on school attendance for boys. Children receiving the shoes were significantly more likely to state that outsiders should provide for the needs of their family. Thus, in a context where most children already own at least one pair of shoes, the overall impact of the shoe donation program appears to be negligible, illustrating the importance of more careful targeting of in-kind donation programs.

This paper is a product of the Operations and Strategy Team, Development Economics Vice Presidency. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at wydick@usfca.edu.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Shoeing the Children: The Impact of the TOMS Shoe Donation Program in Rural El Salvador

Bruce Wydick, Elizabeth Katz, Flor Calvo, Felipe Gutierrez, and Brendan Janet

JEL codes: O12, O15, I31, I32

Keywords: education, health, impact evaluations, in-kind donations, shoes, randomized trial, time allocation

Bruce Wydick (corresponding author) is a professor in the Department of Economics at the University of San Francisco and Research Affiliate, Kellogg Institute of International Studies, University of Notre Dame; his email address is wydick@usfca.edu. Elizabeth Katz is an associate professor in the Department of Economics at the University of San Francisco; her email address is egkatz@usfca.edu. Flor Calvo is a research assistant at 3ie in Washington, DC; her email address is facalvo@dons.usfca.edu. Felipe Gutierrez is a physician with the Veterans Affairs Health Care System and the University of Arizona College of Medicine in Phoenix, Arizona; his email address is felipe.gutierrez@va.gov. Brendan Janet is an independent field research consultant; his email address is bjanet0@gmail.com. We would like to thank Zachary Intemann, Nicole Shevloff, and the directors and staff of World Vision El Salvador for extensive help during field research and David McKenzie, Robert Jensen, and Yaniv Stopnitsky for helpful comments. The insights and suggestions of the three peer reviewers greatly clarified and improved the manuscript. We are grateful to the University of San Francisco's Masters of Science program in International and Development Economics for field support, the USF Faculty Development Fund, and to TOMS Shoes for support and funding of fieldwork.

In-kind donations have become an increasingly common source of direct aid to poor households in developing countries. In many instances, in-kind donations are given through government aid programs, but gifts such as animals, grain, laptop computers, books, clothing, and shoes are commonly provided by private donors in wealthy countries through nonprofit organizations. An important and growing debate exists over the impact of these in-kind transfers, an empirical debate that is now often waged through randomized trials and rigorous quasi-experimental methods to assess the extent to which different types of in-kind transfers exhibit positive effects on beneficiaries.

A series of papers has started to examine the impact of in-kind goods such as used apparel donations in Africa (Frazer 2008), the donation of school uniforms (Evans et al. 2009; Hidalgo et al. 2013), the One Laptop Per Child program (Cristia et al. 2012), the nutritional impacts of dairy cows and meat goats donated through the Heifer Project (Rawlins et al. 2014), wheelchairs for the disabled (Grider and Wydick, 2016) and improved cook stoves (Ludwinski et al. 2011; Burwen and Levine 2012). Related research such as Amed et al. (2009), Cunha et al. (2011), Cunha (2014), and Blattman and Niehaus (2014) compares the causal impacts and cost-effectiveness of in-kind transfers with the impact of direct cash transfers.

Our research contributes to this debate by assessing the impacts of the well-known TOMS Shoes donation program on children’s school attendance, health outcomes, self-esteem, and aid dependency in rural El Salvador. TOMS Shoes was founded in 2006 by Blake Mycoskie and is a for-profit business that incorporates social objectives: when a consumer buys a pair of TOMS shoes, the company gives a pair of shoes to a child in a developing country. The growth of the company has been astounding by any standard: as of early 2016, TOMS had donated more than 50 million pairs of new shoes to low-income children overseas.¹ Moreover, the success of TOMS has spawned an industry of imitators,

1. More recently, the company has begun to sell four other items for which they match consumer purchases with an in-kind donation to an individual in a developing country: (1) sunglasses, where the company carries out vision correction for a sight-impaired individual in a developing country; (2) roasted coffee, where for every purchase TOMS provides a week’s

where the most obvious is “BOBS Shoes” from Sketchers, which has donated ten million shoes to children worldwide, likewise based on an equal number of consumer purchases. But an army of other socially conscious firms have joined the movement: Warby Parker has made “ultra-affordable” corrective eyeglasses available to one million of the world’s visually impaired from an equal number of domestic consumer purchases. Nouri Bar has donated 90,000 nutritious meals to hungry children overseas, one for every nutrition bar sold. Partnering with giving charities such as Global Aid Network and Soles4Souls, for every pair of boots it sells, Roma Boots distributes boots in 20 countries under the mantra of “Giving Poverty the Boot.” The clothing company Uniform has recently launched a Kickstarter campaign to raise funds for a project which would provide a (locally produced) school uniform to a child in Liberia with every purchased item. For each soap product it sells, Soapbox Soaps donates a month of water, a bar of soap, or a year’s worth of vitamins, and the condom company Sir Richard’s boasts of selling—and in turn donating—over three million condoms. The commercial success of these companies has spurred dramatic growth in overseas apparel donations, which are already valued by the Commerce Department at over \$700 million per year.

Before a discussion of impact, a primary question relates to the astounding growth of in-kind transfers: why should an altruistic donor provide an in-kind good that arguably yields a more restrictive outcome than a cash transfer? Currie and Gahvari (2008) provide an excellent review of the motivations for in-kind transfers generally. We supplement their insights by suggesting seven rationales for why donors in wealthy countries might prefer this kind of in-kind giving over cash transfers to beneficiaries in developing countries: (1) a donor believes that a particular in-kind good will bring the recipient greater long-term benefits than would a cash equivalent; (2) in-kind giving may solve a selection problem: to the extent that transaction costs, psychic costs, or social stigma impose relatively heavier burdens on the least needy, then only the most needy in the population may accept the in-kind gift; (3) in-kind gifts may also solve an agency problem—because the gifts are in-

worth of fresh water for an individual; (3) handbags, resulting in the provision of maternal services during childbirth in places in the world where these services are scarce; and (4) backpacks, which fund anti-bullying programs.

kind, donors need not fear that cash is being misallocated in ways that are incongruent with the donor's interests (e.g., spent on "temptation goods" such as alcohol, tobacco, gambling, or prostitution); (4) the nature of the in-kind gift (for example, animal donation or child sponsorship) may be attractive as a marketing tool and is hence more advantageous for fundraising than a cash transfer; (5) donated in-kind goods are in "surplus" such that the value of the goods is exceeded by the tax write-off, warm glow effect of giving, or positive publicity from their donation; (6) donors may have preferences for equality of ownership across particular goods (e.g., healthcare, food, shoes) between the rich and the poor, but not over other goods which could be purchased with cash; (7) some in-kind gifts may yield positive externalities over a community, for example, in the areas of water, sanitation, health, or education, such that private expenditures on the good (resulting from a cash transfer) are sub-optimal.²

In the case of child-specific goods, one related goal may be to encourage poor families to invest in their children's human capital by alleviating the private costs of education and healthcare. Shoes, for example, are frequently a requirement for children to attend school, and they also may play a role in preventing soil transmitted diseases, protecting children from injuries impacting overall foot health, and perhaps impacting overall health as well. While there is no comprehensive cross-country dataset on shoe ownership rates among children, DHS data available from eight other developing countries indicate a wide range of shoe ownership, from 37.1% in Uganda to 97.8% in Vietnam.³ In El Salvador, a nationally representative household survey conducted at the time of this study indicates that 44.5% of all rural households had purchased shoes during the past month, with an average monthly expenditure of US\$4.76 (DIGESTYC 2013).

2. Judging by their corporate communications, for TOMS Shoes, the most important of these rationales would appear to be (1) and (7), where their website states: "What your purchase supports: (1) Improved health, (2) Access to education, (3) Confidence building. Our Giving Partners provide health, education and community development programs to help improve the future of children, their families and communities in need."

3. Other countries for which there is data indicate shoe ownership to be 67.0% in Cambodia, 87.6% in Cote D'Ivoire, 55% and 81% of rural and urban children (respectively) in Namibia, 73.2% in Swaziland, 56.2% in Tanzania, and 62.3% in Zimbabwe. Several donor websites claim that there are 300 million children worldwide without shoes, but it is unclear where this figure originates.

The debate around the efficacy of in-kind international development aid has focused on the impact of such donations on local markets for competing products. For example, a comprehensive review of the existing empirical evidence on the unintended consequences of food aid (by far the largest category of in-kind international aid) concludes that

Although food aid can have negative unintended consequences, the empirical evidence is thin and often contradictory. The available evidence suggests that harmful effects are most likely to occur when food aid arrives or is purchased at the wrong time, when food aid distribution is not well targeted to the most food insecure households, and when the local market is relatively poorly integrated with broader national, regional and global markets (Barrett 2006).

Easterly and Pfutze (2008), in their analysis of the best and worst practices in foreign aid, are less equivocal:

[Food aid] consists mostly of in-kind provision of foods by the donor country, which could almost always be purchased much cheaper locally. Food aid is essentially a way to for high-income countries to dump their excess agricultural production on markets in low-income countries.

The work of Cunha et al. (2011) suggests that in-kind donations reduce domestic prices, finding that the negative price effects of in-kind food donations increased the benefit to consumers by a full 11% of the direct value of the donation, while the impact on local food producers was equivalently adverse. Frazer (2008) finds similar negative impacts of this type on local markets in apparel production in a multi-country study. Looking at cross-country data on used-clothing imports across African countries, he finds that these imports explain roughly 40% of the decline in production in the region and 50% of the decline in employment over the period 1981–2000. However, in a companion paper to this one, Wydick, Janet, and Katz (2014) find no statistically significant impacts on local shoe markets from shoe donations in El Salvador, although regression estimates point to slight reductions in market shoe purchases from households randomly given a pair of children’s shoes. Point estimates indicate a sales reduction of one pair of local shoes for about every 20 pairs of donated shoes.

A recent review by Gentilini (2014) of 12 impact evaluations deliberately comparing cash versus food transfers finds that differences in effectiveness vary by indicator, although they tend to be moderate on average. In some cases differences are more marked (i.e., food consumption and calorie availability), but in most instances they are not statistically significant. In general, transfers' performance and their difference seem a function of interactions among factors like the profile and 'initial conditions' of beneficiaries, the capacity of local markets, and program objectives and design. For example, Cunha (2014) finds that in comparing in-kind food aid with cash grants in Mexico that the small differences in the nutritional intake of women and children under in-kind transfers did not lead to significantly greater improvements in health outcomes when compared to the cash transfers. There is also evidence that recipients tend to use cash transfers mainly for important household expenditures and business investment. In an evaluation of GiveDirectly, Haushofer and Shapiro (forthcoming) find in a randomized trial involving over 1,000 households in western Kenya, no increase in the fraction of expenditures in the treated households on cigarettes, alcohol, or gambling, and that recipients of the transfers spent cash largely on building up business and herd assets (58% increase over the control group mean) and on food, healthcare, education, and social or family events such as weddings and funerals. A comprehensive review of 19 studies on cash transfers from Asia, Africa, and Latin America finds that, in all but two of these studies, recipients do not use cash transfers to increase expenditures on temptation goods (Evans and Popova 2014).

With regard to educational inputs, the evidence appears to vary by region. In Kenya, providing free school uniforms to children reduced absenteeism by 44% and raised test scores by 0.25 standard deviations (Evans et al. 2009). Another Kenyan study focused on adolescent girls also found that providing free uniforms reduced school dropout, teen childbearing, and early marriage (Duflo et al. 2015). However, in Ecuador, attendance actually declined by 25% in the schools that received free uniforms, which may be attributable to the fact that there is not a binding credit constraint that would prevent poor urban households from purchasing school uniforms at market cost, and parents who pay for the children's school uniforms may feel more committed to the school than parents whose children

get the uniforms for free (Hidalgo et al. 2013). Providing subsidized or free meals to schoolchildren has also been found to increase attendance, test scores, and health outcomes in both developing and developed countries (Drèze and Kingdon 2001; Afridi 2011; Gundersen et al. 2012; Vermeersch and Kremer 2005).

Many critics of in-kind transfers contend that allocation of free goods reduces impact. In the health sector, the appropriate transfer mechanism of insecticide-treated bed nets—a good with substantial positive externalities across communities—is hotly debated with some advocating for free distribution and others for some degree of cost sharing. Indeed Warby Parker has used this argument to favor a small charge for their eyeglasses provision overseas in lieu of outright donation. While there are logical arguments in favor of charging a positive price for these highly effective malaria prevention products, experimental evidence from Uganda and Kenya finds no evidence that cost-sharing reduces wastage on those who will not use the product or induces selection of individuals who need the net more, or that households who receive free bed nets resell the donated product to wealthier families (Cohen and Dupas 2010; Hoffman et al. 2009). In the case of shoes for children, whether outright donation is the most effective way to achieve the desired impacts is likely to be highly sensitive to the context, including existing beliefs and practices and the market availability of footwear.

Our study on the impact of TOMS in-kind shoe donations was carried out in El Salvador, a lower middle-income country (GNI per capita \$3,590) with a population of 6.3 million. We measure impacts from in-kind shoe donations based on the results of a randomized trial involving 1,578 children across 18 rural communities. Each of these 18 communities was located in a World Vision Area Development Program (ADP) region that consisted of four to six communities, half of which were randomly selected for treatment and half for control. By random chance, some of the larger communities within the ADP regions were selected into treatment so that our study contains 912 children in treated communities and 666 in control. In the nine treatment communities, children received a pair of donated canvas loafers at baseline (see figure 1), while children in the other nine communities served as controls. Our choice was a cluster randomization at the community level in order to address human subject concerns

related to jealousies that might arise from randomizing the shoe gifts at the household level within villages. Because of the cluster randomization and the likelihood of intraclass correlation within communities, it was important to obtain both baseline data before treatment and follow-up data on our subjects, and thus we use difference-in-difference and ANCOVA estimations to analyze the impacts of treatment, but still a relatively small number of clusters does reduce statistical power. To adjust for our relatively small number of clusters, we use wild-bootstrapped standard errors in ascertaining statistical significance in most of our estimations.

We obtain data on a significant number of outcomes for children potentially affected by wearing shoes: time allocation across 12 different children's activities, impacts on school attendance (since shoes are required to attend school in El Salvador), impacts on general health and foot health specifically, impacts on children's self-esteem, and effects on aid dependency among children.

Our results show heavy usage of the donated TOMS shoes; indeed, the modal response was that children in the treatment group wore the shoes every day of the week. However, our study finds that the donated shoes did not significantly reduce shoelessness and had insignificant impacts in most of our key categories that would indicate transformative impacts. There are two likely reasons for this lack of impact. The first is related to targeting: in a middle-income country such as El Salvador, where the majority of children already have shoes (either purchased or donated from the government), an additional pair of shoes is unlikely to have first-order effects such as improved health or schooling, stated goals of the donor.⁴ A second reason for low impact is that the quality differential and monetary value of the shoes (approximately \$3–\$5) is probably too low to bring about detectable second-order income effects (for example, by freeing up cash for other expenses).

While we find modest evidence in our ANCOVA estimates for small positive impacts on school attendance, we also find some evidence of negative impacts in the area of aid dependency, though this

4. Of particular importance with respect to the hypothesized effect of shoe ownership on school attendance is the recent implementation of a national program to provide school uniforms, shoes, and supplies to all primary school students in El Salvador. Moreover, many children in the communities included in the study have received benefits from World Vision's child sponsorship program, including previous distributions of free shoes.

evidence for aid dependency is not behaviorally based but strictly attitudinal. Our conclusions from the study closely mirror those of Barrett (2006) and emphasize the importance of thoughtful and careful targeting of in-kind donations and the manner in which in-kind gifts are allocated to beneficiaries. To the extent that in-kind giving is able to realize significant enhancements in welfare compared to cash, it must be targeted at communities in which the in-kind good is scarce, but nevertheless exhibit a high demand for the good. And these instances may be less commonplace than many donating organizations are willing to admit. To reduce negative psychological impacts related to aid dependency, in-kind goods such as shoes can be given as incentives for undertaking other positive behaviors, such as obtaining vaccinations, health examinations, and school attendance or performance. Given as incentives, beneficiaries may be more likely to associate the gift as a natural reward to their own positive effort, perhaps mitigating feelings of aid dependency or entitlement. Thus, it appears there may be a number of valid circumstances in which in-kind giving is optimal when carried out carefully. But absent these conditions, conditional or unconditional cash transfers may be more appropriate.

The remainder of the paper is organized as follows: Section 2 provides background and context for the study, including the results of moderated focus groups carried out prior to the field experiment. Section 3 describes the data, hypotheses, and empirical model. In section 4, we present the results of the experiment, including an impact narrative of the median impact subject of the treatment group in the TOMS experiment. Section 5 concludes with a discussion of the programmatic implications of the research.

I. FIELDWORK SITE AND BACKGROUND

Although it is a relatively small country, economic development is uneven across El Salvador. Therefore, in conjunction with TOMS' giving partner, World Vision International, our community-level selection criteria for inclusion in the study were oriented toward choosing the poorest communities in the country and those having the highest levels of shoelessness among young children. Specifically, utilizing the poverty map constructed by the *Dirección General de Estadística y Censos*

(*DIGESTYC*) based on nationally representative household survey data, municipalities with high and severe levels of income poverty were matched with four subregions (Area Development Programs or ADPs) where the implementing nongovernmental organization had ongoing operations. ADPs that had carried out shoe donation activities in the past year were excluded. Secondary selection criteria included safety/accessibility for surveyors and cooperation with local schools and shoe vendors. The four selected ADPs were in the municipalities of San Julian (M9 in figure 2), Ozatlan (A36), San Francisco de Javier (A7), and Carolina (S10).

Our study took place three years after a social democratic government had come into power, making a significant commitment to poverty alleviation and social inclusion, devoting additional public revenues to investments in health and education, and undertaking a comprehensive program of education reform (Mills 2012). As part of these reforms, the *Paquetes Escolares* program began to provide uniforms, school shoes, and school supplies like notebooks, textbooks, and writing supplies to children in grades K-9. The items distributed by this program utilize local and national materials and vendors. Coverage has been broad: between 2011 and 2012, 1,386,767 students received all or part of the “package,” which includes two uniforms, one pair of shoes, and one set of school supplies, so that virtually none of the children in our study (specifically only two) were completely shoeless at baseline.

As part of the preparation for the design of the experiment and survey instrument, we held focus group discussions in four communities that had been previous beneficiaries of shoe donations. The focus group protocol contained 45 guided questions, and discussions carried out with groups of 10–20 mothers in each community. Four principal topics were addressed in the focus groups: children’s shoe wearing practices and market access to shoes; past experience with shoe donation programs; health and sickness; and time use patterns among children.

Current Shoe Usage

Focus group discussions conducted before our experiment revealed that, in three out of the four communities, a majority of the children did not wear shoes for much of the time outside of school. When they did wear shoes, children usually started to wear them between one and two years of age. In

general, children wore all types of shoes, including leather shoes for school, canvas tennis shoes, sandals, Crocs (plastic clogs), and rubber flip-flops. While the response from the mothers was that children wore all types of shoes, the observation among the research team was that the majority of children wore sandals, Crocs, or nothing at all. Children generally wore their nicest pair of shoes for special events such as church, parties, or community gatherings (such as the focus group discussion). Besides special events, most children were only required to wear shoes at school. Focus group participants revealed little difference in the type of shoe worn by female versus male children.

Shoes were purchased from various locations, including a smaller urban center located five to 45 minutes away on foot from most communities and a larger city located farther away but accessible to the communities by bus. Shoe prices varied depending on the type of shoe, with cheaper sandals and crocs averaging three to five dollars and nicer shoes costing \$20 or more.⁵ Shoes are required for school attendance in El Salvador, and many of the public schools were reported to distribute the black leather shoes at no cost to the students as part of the *Paquetes Escolares* program. This program, however, does not reach every school in El Salvador, and there was a significant delay in receiving the shoes (children receive shoes one year after they are registered in school). The consensus in focus groups was that children often preferred to be shoeless. This appeared to be somewhat a product of custom, but also because parents appeared to put pressure on the children to take care of the shoes they had, and thus children often did not wear shoes in order to preserve the one nice pair they owned.

Experience with Shoe Donation Programs

In late 2011, TOMS began donating shoes through its giving partner World Vision in many of the communities around our study area. Most of the children who received the shoe donation already owned a pair of sandals or shoes (some of which were the free pair they had received from school). The donated TOMS shoes were generally popular, but not universally so. Some in the focus groups reported that boys in the communities thought the shoes were only for girls, while others joked that

5. The market value of the donated shoes is on the lower end of this spectrum (\$3–\$5).

the shoes “looked like they were for pregnant women.” Others also complained that the shoes were not truly unisex as claimed. The differences between purchased shoes and donated shoes were evident among the population. The mothers liked having the free shoes for their children but often indicated they would have preferred shoes with laces (sneakers).

Usage of the previously donated shoes varied between the communities. Two of the focus group communities reported high usage of the shoes, while others reported very little usage of the shoes. Children used the previously donated shoes for playing soccer, going to school, going to church, and community meetings. They commented that the shoes were not very durable and usually didn’t last long enough to be passed down to a younger sibling.

Health Environment for Children

Mothers understood the importance of wearing shoes: to protect feet from cuts, sprains, parasites, and fungi. They reported that the most common illness among children were parasitic infections, such as *amebiasis*, typically transmitted via fecal-oral route and contaminated water supplies. Common symptoms were vomiting, diarrhea, abdominal pain, poor appetite, and fever. With regard to foot health in particular, the focus group discussions revealed two main concerns among local children: foot injuries during hard play and chores and fungal foot infections (particularly during the rainy season), the latter of which can cause breakdown of skin as a barrier to infection, rashes, itchiness, and malodor. While there is no single standard defining foot health, we focused on the foot health issues commonly encountered in the participating communities.

A pilot study to establish the baseline prevalence of soil transmitted parasites with an emphasis on hookworm in rural El Salvador revealed that, while intestinal parasitic diseases were present, protozoan organisms predominated. In fact, we found zero evidence of soil transmitted helminth infections. Thus, it appeared that a government effort to control soil transmitted diseases by large scale distribution of albendazole had succeeded, even reaching the most remote parts of the country. As a result, our study context was not appropriate for carrying out an impact study of shoes on the prevention of hookworm infection.

To gauge schooling and time use patterns of children in rural El Salvador, we asked mothers about school attendance and the typical nonschool activities of school-aged children. A majority of the children were reported to be enrolled and attended school regularly, and there was a perception in the focus groups that school attendance rates had risen due to the conditional cash transfer program in El Salvador, *Comunidades Solidarias Rurales*, that provided \$15 a month for a child attending school.

In these communities, children walked to school, and walking travel time was typically between 15 and 30 minutes to school. Children began attending primary school at the age of seven and finished primary school around age ten or eleven. The younger children attended school from 7:30 a.m. until 1:00 p.m., and the older children attended school from 12:30 p.m. until 5:00 p.m.

Mothers reported that children typically awoke at about 6:00 a.m., showered, ate breakfast, and prepared themselves for school. After school, children then typically ate lunch and engaged in a number of different activities such as playing soccer, watching television, and doing homework and chores. Chores that boys were reported to carry out most often were helping with the maize harvest, grazing cattle, gathering firewood, sweeping the house, and going to the grain mill. Chores for girls most often included washing clothes, going to the grain mill, sweeping the house, running errands, caring for younger siblings, and collecting the trash. Most of the time, children were reported to take off their school shoes upon arriving home and then put on a pair of sandals or simply go barefoot.

II. HYPOTHESES AND EMPIRICAL MODEL

After carrying out focus group research, we randomly selected study communities in each of the study regions in which shoes had not yet been donated. Within each of the four ADPs, four to six communities in the ADP were selected for the study based on the presence of local schools and lack of shoe donations in the past 6–12 months. Communities within a given ADP were very similar to one another; communities in different ADPs were often quite different culturally and geographically. For this reason, the randomization took place within each ADP. From the four to six communities selected in each ADP, half were randomly assigned treatment status, and the other half served as controls.

Community meetings were held at the local schools to announce the study and collect home addresses from the participants. Children in treatment communities received shoes immediately after the baseline survey, while age-eligible children in the control communities received shoes three months later after the follow-up survey. From each community, households with children seven to 12 years old constituted the target population from which the sampling frame was constructed. An average of 367 children from each of the four ADPs was included in the study.

Survey

Our baseline survey undertaken from July to October of 2012 included questions on family structure, health, foot health, education and schooling, shoe purchases, migration, time allocation, and land and asset ownership. The follow-up survey was carried out November 2012 to February 2013, three to four months after the baseline survey and disbursement of shoes and also included psychological questions on self-esteem and aid dependence. Enumerators conducted the survey at the household with the adult (usually mother) and child; interviews lasted approximately one hour.

Time Use Diary

We randomly chose half of our sample of children for a time-use survey. Attached to each of these surveys was a one-page time-use diary (see supplemental appendix S1). Parents were informed during the pre-survey meeting that they should carefully observe their child's time use on the day prior to participating in the survey and make notes if possible. Along the vertical axis, the diary included categorical pictures of all possible activities that a child performs in a 24-hour period (school, eating, play, homework, etc.). Along the horizontal axis, the diary included one-hour time slots from 6:00 a.m. to 9:00 p.m. (assuming the child was sleeping the rest of the night). Within these timeslots, the enumerator could make a checkmark under the appropriate activity in which the child was engaged. The recall period was the previous 24-hour period, with the exception of interviews conducted on Mondays, which referred to the previous Friday in order to capture school attendance.

Shoe Donations

Shoe donations were carried out in the treatment communities immediately after the first round of surveying. Donations took place at a central location in each community, usually the local school. All children between the ages of seven and 12 years old in treatment communities whose families participated in the survey were carefully fitted with a black pair of TOMS' canvas, rubber-soled loafers. Shoe fitting and donation was carried out in the control communities just after the second round of surveying was complete. Those in the control communities were unaware that they would receive shoe donations in the future.

Chain of Causal Impacts and Registered Hypotheses

When donors provide in-kind aid, they frequently have an implicit or explicit theory of change with respect to the intended impacts from donations. A theory of change “depicts a sequence of events leading to outcomes, explores the conditions and assumptions needed for the change to take place, makes explicit the causal logic behind the program, and maps the program interventions along logical causal pathways” (Gertler et al. 2011). A theory of change associated with the shoe donation program can be seen in a results chain, mapping inputs, activities, outputs, and outcomes of the intervention. This conceptual framework implies the following set of potential causal effects along the chain:

1. Shoes are effectively fit and distributed to children.
2. Children own an added pair of shoes, wear them, and the time they are shoeless is reduced.
3. Children's time use is altered in favor of activities requiring or benefiting from shoe-wearing.
4. School attendance and outdoor play increase with corresponding benefits to children.
5. Children exhibit higher levels of self-esteem from ownership of the shoes and from enhanced participation in esteem-building activities.

Based on this framework, we developed a series of hypotheses that we registered in a pre-analysis plan at the JPAL hypothesis registry.⁶ All hypotheses and regression specifications were submitted before examination or analysis of any of our data and include our study on the market impacts of the donations as well as their impacts on children. The full pre-analysis plan is included in supplemental appendix S2 and can be viewed online on the JPAL website at the URL included in the appendix.⁷

The corresponding null and alternative hypotheses as specifically laid out in our JPAL hypothesis registry were the following:

i) H_0/H_a : No impact (positive impact) of receiving donated shoes on child school attendance.

ii) H_0/H_a : No impact (positive impact) of receiving donated shoes on children's allocation of time toward activities that are facilitated by shoe-wearing. For example, our alternative hypotheses would suggest that children would allocate time away from activities such as watching TV and toward playing sports.

iii) H_0/H_a : No impact (positive impact) of receiving donated shoes on children's foot health.

iv) H_0/H_a : No impact (positive impact) of receiving donated shoes on children's self-esteem and psychology.

v) H_0/H_a : No impact (negative impact) of receiving donated shoes on children's sense that families should provide for their own needs rather than having others provide for them.

Our pre-analysis plan specified the corresponding empirical model to be estimated for (*i*), (*ii*), and (*iii*), namely

$$y_{ijt} = \alpha + \mathbf{X}_{ij}'\boldsymbol{\beta} + \theta T + \tau TF + \pi F + \mu_j + \varepsilon_{it} . \quad (1)$$

6. Our registry can be viewed online at <http://www.povertyactionlab.org/Hypothesis-Registry>. Casey et al. (2012) provide a rationale and framework for pre-analysis plans.

7. Note that, as of 2013, JPAL is no longer accepting pre-analysis plans for hypothesis registry, instead deferring hypothesis registry to the American Economic Association site.

Because our psychology data in (iv) and (v) are solely taken at endline, for these we use a post-estimator with the following specification

$$y_{ij} = \alpha + \mathbf{X}_{ij}'\boldsymbol{\beta} + \theta T + \mu_j + \varepsilon_{it}, \quad (2)$$

where y_{ij} is the relevant impact indicator, \mathbf{X}_{ij}' are control variables that describe the child and her household characteristics, which will include age, gender (male =1), economic activity of parents, and indices of dwelling quality and asset ownership; T is an indicator of whether the child lives in a treatment community, F denotes an observation in the follow-up period (as opposed to the baseline period), μ_j is an ADP (region)-level fixed effect, and ε_{it} is the error term. Impact is captured by the coefficient on the interaction term, τ , in our difference-in-difference estimations in (1) and by θ in (2).

An important piece of research highlighting the more efficient estimation of ANCOVA relative to differences-in-differences for experimental data (McKenzie 2012) was published only a few months before our hypothesis registry. Because we were unaware of the relative efficiency of ANCOVA in analyzing baseline and follow-up data in field experiments such as ours, we include the ANCOVA estimations in our paper next to the estimations we carry out through difference-in-differences. The ANCOVA combines regression with ANOVA (analysis of variance) and produces estimates of average treatment effects using baseline survey data as a right-hand-side control. In the case of a single baseline and follow-up survey, the specification is

$$y_{ijt+1} = \alpha + \theta T + y_{ijt} + \mathbf{X}_{ijt+1}'\boldsymbol{\beta} + \mu_j + \varepsilon_{it}. \quad (3)$$

McKenzie (2012) builds on Frison and Pocock (1992), which demonstrates that, under reasonable assumptions, the ANCOVA estimator is more efficient (retaining unbiasedness with lower variance) than both the post-estimator and differences-in-differences with experimental data. Indeed, the empirical results we present from our TOMS experiment show smaller standard errors relative to our difference-in-difference estimates though smaller point estimates as well.

To address the issue of over-testing, we create an Anderson Index (see Anderson 2008) within each of our variable families (e.g., Health, Foot Health, Self-Esteem) as stipulated in our pre-analysis plan. The Anderson Index is created by orienting variables in a single direction of impact, de-meaning and

normalizing each of the dependent variables in the respective group j . Differing from the more common index of Kling et al. (2007), which calculates a simple average of normalized variables, the Anderson Index assigns a weight on each impact variable by the sum of its row entries across the inverted variance-covariance matrix of the impact variables in the group j . Specifically, each variable i in group j receives a weight, or index score, of $\bar{s}_{ij} = (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}(\mathbf{1}'\Sigma^{-1}\mathbf{y}_{ij})$, where $\mathbf{1}$ is a $m \times 1$ column vector of 1's, Σ^{-1} is the $m \times m$ inverted covariance matrix, and \mathbf{y}_{ij} is the $m \times 1$ vector of outcomes for individual i . The Anderson Index assigns weights to variables such that a variable within the family that exhibits lower covariance with the other variables becomes weighted proportionally higher in the index because it contains more independent information. This form of indexing allows us to create more general conclusions about the impact of the shoes on a particular family of outcomes, and helps address the issue of over-testing that could erroneously assign too much importance to a possibly spurious rejection of a single null hypothesis for one variable within a family of outcomes.

III. Empirical Results

Table 1 shows balancing tests between our treatment and control communities. The average age of children in the study is 9.33 in treatment relative to 9.49 in control. The proportion of boys is slightly higher in control communities, 54.5% versus 49.7% in treatment. Households in the treated communities are little more likely to work in agriculture, and their parents are slightly less educated, 5.35 years versus 5.84 years in control, although the dwelling quality index and consumer durable indices are a little bit higher in the treatment areas. Among those chosen for the time survey, children in the treated group allocate significantly more time at baseline to activities such as shopping, church, fetching water and firewood, and working outside the home, but less time sleeping, in school, and in household chores, noting that these raw values do not account for seasonality at time of survey as do our regression estimations. Table 1 shows baseline hours of shoe ownership and shoelessness during waking hours are quite similar, 2.06 and 1.82 pairs and 2.09 and 1.96 hours among children in treatment and control, respectively. Children in treated communities rank (insignificantly) lower in the

health index, but (significantly) higher in the foot health index at baseline. We include 25 control and impact variables in our balancing test and find four of these variables to be significantly different (at just the 10% level) in simple cluster-adjusted t -tests. Most of these, however, are variables related to our time-allocation study, and without controls adjusting for time of year of the survey, it is unsurprising that we find differences here at baseline between treatment and control.

Difference-in-Difference and ANCOVA Estimation

Our regression estimations in tables 2–8 examine the impact of the TOMS shoe donations on children’s time allocation, school attendance, the health of children’s feet and their general health, on measures of their self-esteem, and on aid dependency. In each of our estimations we include ADP-level fixed effects and cluster our standard errors at the community level as stipulated in our pre-analysis plan. Because our randomization was carried out over a relatively few number of clusters (18), except for our SUR estimations we estimate clustered standard errors using the wild bootstrap method of Wu (1986) and Cameron, Miller, and Gelbach (2008).⁸ Because of the relatively low intra-class correlation in our sample, the standard errors obtained from the wild bootstrap (1000 iterations) are only marginally bigger than conventional clustered standard errors in most estimations.

In keeping with our causal chain, hypotheses, and our empirical model, we first check whether shoe ownership increased and shoelessness decreased. This is not a foregone conclusion as it might be possible for shoe donations to substitute for shoe purchases, where saved resources could be allocated to other goods. At baseline, children in the treated communities owned an average of 1.82 pairs of shoes. After the shoes were given away, this increased significantly to 2.31 pairs per child. However, average shoe ownership also increased in the control communities by 0.26 pairs, such that difference-in-difference estimations do not find significant increases in shoe ownership. Table 2 presents regression results on the first step of our causal chain, whether the shoe donation intervention increased shoe ownership and reduced shoelessness. These estimates show even smaller increases in

8. Angrist and Pischke (2009) suggest that the use of the wild bootstrap or other corrective measures are appropriate for data analysis where the number of clusters is smaller than 42.

shoe ownership, where point estimates find that shoe ownership increased by only 0.22 pairs of shoes (difference-in-differences) and 0.075 (ANCOVA) respectively. Neither do we find that shoelessness time among children in the treatment group was reduced. From nearly identical baseline levels, average hours of shoelessness fell by 0.39 hours in the treatment group, but by twice as much in the control group (0.76 hours).

Table 2 gives difference-in-difference and ANCOVA regression estimates that include controls and also do not yield evidence for reduction in shoelessness. In this table as in tables 3–7, the parameter estimate of interest in difference-in-difference estimations for the purposes of program impact evaluation is the coefficient on *Treated* \times *Round 2*, which shows the differential change in the dependent variable over the treatment period across children in communities that received shoe donations and those that did not. For ANCOVA estimates, which control for baseline realizations of the impact variables, the coefficient of interest is *Treated*. These display positive (but insignificant) coefficients of 0.40 hours and 0.64 hours, respectively.

Children did, however, wear the donated shoes. Our follow-up survey found that just over 90% of children wore the new shoes, and almost 77% wore them at least three days a week. Figure 3 provides a histogram of the intensity of wearing the shoes by days per week. Indeed, the modal response was that a treated child actually wore the donated TOMS shoes seven days a week. Looking at the donated shoe use by activity, children appear to have worn them most frequently as house shoes, for play, and for school (see figure 4). The data therefore do not show any significant social or attitudinal barriers to footwear use, as has been found in other settings (e.g., Ayode et al. 2013). Our endline survey also found that 95% of the treated children had a favorable impression of the shoes.

While it is clear that children liked and wore the donated TOMS shoes, the question remains whether the donated shoes simply substituted for older shoes. Although we find no evidence that the shoe donations reduced shoelessness among the children in the treatment group, it is thus possible that the shoes substituted for shoes of lower quality and thus still realized positive impacts on our outcome variables. Therefore, we examine the third, fourth, and fifth links in the results chain: do the shoe

donations alter the children’s time allocation, yield increases in health and school attendance, and thus perhaps realize significant changes in their psychological outcomes and self-esteem?

The time allocation equations were estimated using Seemingly Unrelated Regressions (SUR) due to strong correlation among the error terms among the twelve categories of time use, since time allocation must sum to a constant.⁹ It is important in our time allocation estimates to account for seasonality due to changes in activities across cropping seasons, seasonal weather patterns, and schooling seasons. This is especially critical because due to logistical constraints approximately 65% of our second-round surveys took place during months when children were not normally attending school. To account for these expected seasonal differences, our regressions include dummy variables for each month of the year (omitting January) during which the baseline and follow-up time data was taken. We find relatively minor impacts from the shoes on time allocation, but again there are hints that the shoes moved time allocation away from indoor activity to outdoor activity. The coefficients in table 3 can be interpreted as changes in fractions of an hour in the relevant activity. The most notable point estimate changes are a difference-in-difference estimated reduction of nearly half an hour per day watching TV (not significant), and a decrease in time spent doing homework of 0.28 of an hour ($p < 0.05$). However, while the ANCOVA estimates carry the same sign, they are much smaller and have lower variance. While our simple endline data shows an increase in outdoor activity of 1.02 hours per day, both regression estimates are insignificant.

In table 4, we find modest evidence that the shoe donations had an impact on school attendance. These regressions use school administrative attendance data gathered from the local primary school at baseline and follow-up. Because of the large fraction of surveys taken after the end of the school year, we used administrative attendance data for the last school month of the year (October). We look at the impact on attendance overall and break our sample into impacts on boys and on girls. Difference-in-difference estimates in column 1 in table 4 indicate that the children in the treatment group missed an

9. Due to the difficulty of obtaining standard errors using the wild-bootstrap method in a SUR framework that exploits correlation between error terms across equations, we use conventional clustered standard errors in our SUR estimations.

average of 0.29 fewer days of school per month (from a baseline mean of 0.70 days) than children in the control communities, but the point estimate is statistically insignificant. ANCOVA estimates, however, indicate a larger impact and are also more precisely estimated. The ANCOVA estimate in column 1 finds a significant reduction in absenteeism of 0.165 days per child per month in the treatment group, significant at the 5% level of confidence. As seen in column 2, most of this impact is on boys, where there is a reduction of 0.195 days, whereas on girls the impact is smaller, although we want to be clear to note that breaking the schooling impacts apart by gender was not part of our pre-analysis plan. It is possible that some of the TOMS shoes substituted for lost or worn school shoes. However, we hesitate to place too great a weight on these results because of their relatively small magnitude, but also because of the *Paquetes Escolares* program, whose goal it was to provide school shoes for all children. The donated TOMS shoes, in contrast, were more frequently worn for home use than to school (see figure 4), and so in this context it is surprising to find positive impacts on school attendance. Moreover, we fail to find reductions in general shoelessness, increases in foot health and general health that would seem to accompany a large and significant reduction in school absenteeism.

In our health estimates (tables 5 and 6), there are no statistically significant differences in any aspect of foot health, or in the overall foot health index in table 5, although many point estimates lie in the direction that we would not expect, with greater incidence of cuts on the feet, foot infections, skin irritation, missing toenails, blisters, and sores among children receiving the donated shoes. It may be that this is due to a type of compensating differential, in which children engage in more outdoor activities and play harder in these activities because they are wearing shoes instead of playing barefoot, but this is only speculation. Our aggregated foot health index is negative for difference-in-differences and just marginally negative for ANCOVA, indicating worse foot health among children receiving the shoes, but not statistically significant.

Our results in table 6 regarding overall health also indicate insignificant effects from the shoe donations. Point estimates indicate slightly higher rates of body injury, consistent with the idea that children are engaging in more outdoor activities or undertaking these activities more aggressively

with the shoes, but these estimates are insignificant. Other than bodily injury we would expect small impacts on general health outcomes, and this is what we indeed find. The health index indicates somewhat better overall health, but this is statistically insignificant.

Our survey of psychological variables used in tables 7 and 8 were only obtained at endline, and so here our relevant coefficient that we report in these estimations is on *Treated*. The self-esteem results in table 7 show that controlling for age and sex of the child, as well as for several measures of the household's socioeconomic status, children who received shoes report greater feelings of capacity and satisfaction with themselves, but lower levels of pride in themselves. While both of these contradictory findings are significant at no less than the 10% level using wild-bootstrap-estimated standard errors, there is no conclusive evidence that receiving the shoes has a positive or negative impact on the psychology of recipient children. The impact on the self-esteem index is essentially zero. Arguably, given the inframarginal nature of the treatment, it may have been more appropriate to measure subtler subjective impacts of receiving the donated shoes. For example, in their study of the impact of piped water adoption in urban Morocco, Devoto et al. (2012) find significantly increased levels of satisfaction and well-being, even in the absence of health effects.

Obviously, one reason for failure to reject null hypotheses in any study could be lack of statistical power. Evidence of a low statistical power would include small minimum detectible effects (MDEs) and consistently large point estimates that yet fail to reject the null hypothesis of no impact due to large standard errors on the estimates. Our relatively small number of clusters (18) may compound this problem, and as the wild-bootstrap procedure corrects standard errors to account for the number of clusters, the somewhat larger standard errors yielded by the correction could compound the problem. Could it be that finding a lack of impact on general health, foot health, and self-esteem is due to under-powered tests?

While a small number of clusters decreases statistical power, we argue that the randomization over a relatively small number of clusters in our study is unlikely to account for the failure to reject multiple null hypotheses of no impact. This is primarily but not solely because, as table 2 illustrates, the

intervention does not appear to have significantly increased shoe ownership among children, an impact measured with substantial precision. Here our ANCOVA point estimate is that the TOMS intervention increased shoe ownership by 0.075 shoes per child with a (wild bootstrapped) standard error of 0.065). Consider an *ex-post* analysis of minimum detectible effect, using a standard p -value of 0.05 with power set at a 0.80 probability of rejecting a false null, at $2.8\times$ the standard error of the regression coefficient ($1.96 + 0.84$ standard deviations). This yields a minimum detectible effect (MDE) of 0.182 shoes. Thus, if the donation increased shoe ownership by even $1/5$ of a pair of shoes on treated children, our estimation would have been able to reject the null hypothesis of no impact on shoe ownership with over 80% probability. Without an impact on shoe ownership, it is difficult to reason that there would be strong impacts on other variables, although a greater number of clusters would add to the statistical power vis-à-vis our hypotheses. This is accentuated by our point estimate that shoelessness actually *increased* slightly in the treated group. Yet the MDE for other of our key variables such as reduced absenteeism is only 0.22 schooldays. For our child health, our MDE is 0.30 of a standard deviation. Thus, it is far more likely, based on the consistently small point estimates, that failure to reject the null stems from the donated shoes exhibiting little influence on these particular outcome variables because the donated shoes did not significantly reduce shoelessness.

Our additional results on psychological impacts of a dependency in table 8 indicate that children receiving the shoes have a much greater propensity to state that “others should provide for the needs of my family.” Indeed, our point estimate shows an increase in the propensity to answer this question in the affirmative by 12.2 percentage points over a mean among the control of 66.4%. We also check whether the donated shoes caused a reduction in those children who have a strong form of “self-sufficiency,” children who both believe that the family should provide for its own needs AND do not believe outsiders should provide for the family’s needs (control baseline 36.8%). We find that the TOMS intervention reduced this economic self-sufficiency by 12.9% toward some form of stated dependency. It is important to note again that our regressions in tables 7 and 8 use endline data only since the psychology questions were only presented in the follow-up survey. Moreover, these findings

reporting stated beliefs and not measured behaviors. But with these caveats, we find it likely that the TOMS donations created some degree of feelings of dependence on outsiders for aid.

As a check for robustness of our aid dependence finding, we carry out a modified version of Fischer’s exact test. This test is often used in the case of small samples or sometimes in the case of a cluster-randomization where the number of clusters is relatively small. The classical use of this method occurs by carrying out a placebo test to calculate the t -statistic that corresponds to every permutation of possible placebo treatment. In this use of randomization inference, every possible permutation of treatment assignment is considered, calculating t -statics for each possible assignment of treatment and control across groups (ignoring the actual assignment to treatment). Then one ascertains if the t -statistic of the true treatment assignment falls above a critical percentile of t -statistics (creating a pseudo p -value) in this exhaustive set of treatment permutations. In our research design, the number of clusters is larger than typical for this exercise, and hence we carried out a simulation in Stata of 1000 random assignments to treatment, where in each of our four World Vision ADP regions, two to three communities were randomly assigned to treatment. In the simulation of the regression equation in table 8, the true t -statistic of 2.65 corresponding to the 11 percentage point increase in treated children to respond that “others should provide for the needs of my family” lies in the 2nd percentile, identical to the p -value on the treatment coefficient in the original regression.

Taken together, these results tell an unexpected story about the effects of the shoe donations in the context of our study. It appears that, while usage of the shoes among children is very high, and approval of the shoes is also very high, the donated shoes do not exhibit the donor’s expected and intended impacts on reduced shoelessness, foot health, general health, and appear to produce mixed effects on self-esteem and some negative effects on feelings of economic self-sufficiency. Instead, it appears that the shoes may increase a child’s time allocation into outdoor activities, which also may be associated with slightly higher rates of injury. And while children appear to like the shoes, and ownership and use may contribute to a greater senses of accomplishment and self-satisfaction, these

positive psychological effects are counterbalanced by what appear to be negative effects of the donations in terms of creating some form of aid dependency.

The Median Impact Narrative

Poverty organizations, including both TOMS, World Vision, other NGOs, and the many socially conscious businesses that have imitated the TOMS model, commonly use anecdotes and narratives of successful program participants in marketing and fundraising efforts. The standard practice for virtually all nonprofits working in the poverty industry, however, is to carefully hand-pick narratives of successful participants—even positive outliers—who have realized program benefits that significantly exceed the average impact of the program. One reason narrative is often used in marketing instead of data may be a lack of attention to rigorous program evaluation. But perhaps a more important motive is that narrative has been rigorously found to exhibit a much stronger motivator for human action than (even very convincing) data analysis.

Small, Loewenstein, and Slovic (2008), for example, present results from an experiment in which subjects, given the opportunity to donate to Save the Children, contributed far more to an “identifiable” victim described in a short narrative of a young girl in poverty than in response to data that conveyed the greater scope of privation, a “statistical” victim. This and other subsequent research (see for example, Bal and Valkampt [2013] and Hsee et al. [2015]) has demonstrated not only how the human brain tends to absorb and process narrative more effectively than data, but how the brain also translates narrative more effectively into effective action.

Narratives need not present a biased picture of causal impacts. Indeed, both narrative and data can equally present both biased and unbiased pictures of causal effects. In response, Wydick (2015) suggests the use of a “median impact narrative” as a way to create a picture of program impact based on the experience of the individual in a treatment group whose response to treatment most accurately captures a median picture of causal effects on program beneficiaries.

We obtain the median impact narrative in the following manner: Let $\mathbf{\Omega}$ equal the $m \times m$ covariance matrix of the m (dependent) impact variables in a study from the treated group, and let ω_j equal the

sum of the row entries for row j of $\mathbf{\Omega}^{-1}$. We weight each impact variable by $w_j = \frac{\omega_j}{\sum_{j=1}^m \omega_j}$. Letting \mathbf{w}_j equal the $1 \times m$ row vector of these weights and \mathbf{y}_j equal the $m \times 1$ column vector of (endline minus baseline) changes in impact variable j , we create an impact index, $I_i = \mathbf{w}_j \mathbf{y}_{ij}$, which, similar to the Anderson Index, places heavier weight on impact variables containing a greater degree of unique information in the sense that they are not highly correlated with other impact variables. The median impact narrative is taken from the treated individual i in the sample ranking in the 50th percentile of the impact index. (These represent baseline to endline changes in the treatment group; average treatment effects are of course obtained by subtracting from these the average change over time in the control group.) For time allocation, school attendance, and health impacts, we smooth the narrative to include the average impacts in the middle quintile of the treatment group. Based on our impact index, I_i , the median impact from the TOMS Shoes donation in our experiment was realized by an eight-year-old boy in the treatment group, José Mantaro, whose narrative we recount:

“José Mantaro lives in San Francisco de Javier, El Salvador, is the son of a young single mother, and has one younger brother.¹⁰ His mother’s rented house reflects her poor economic situation; the house has a dirt floor with an old corrugated iron roof, and it sits on a 25×25 meter plot of dry land. Their house has no electricity or indoor plumbing. José’s family does not own a refrigerator, television, or radio, but both he and his brother have a bike. A member of the family must walk about a kilometer away to get fresh water. José walks 30 minutes to school, where he is in the third grade, but he has not yet learned to read or write. His mother had not purchased shoes for him in the 6 months prior to our baseline survey, in which she indicated that she could not afford new shoes for him.

“José received a pair of TOMS shoes immediately after the baseline study in June 2012 through his community’s involvement with World Vision. He previously owned two other

10. The name of the median impact subject was a combination of arbitrarily chosen first and last names in our sample to protect confidentiality.

pairs of shoes, his school shoes and an older pair of flip-flops. When he received the donated shoes, his mother finally threw away his old pair of flip-flops. José himself reported in the follow-up survey that he very much liked the TOMS shoes and found them comfortable on his feet. He wore the shoes nearly every day during the study period, typically six days a week. During the four-month test period, José's overall health improved slightly, but not significantly more than the average change in health of the control group of children who did not receive shoes. Changes in José's foot health were a little worse than the group not receiving the shoe donations, but by only 0.10 of a standard deviation. Receiving the shoes may have slightly reduced José's absenteeism from school. He spent a few more minutes more per day working outside the home, while reducing his time spent in household chores by about the same amount of time. He spent a little bit more time collecting water for his family and a little less time per day watching TV. An honest appraisal would suggest that receiving the shoes did not bring about transformative changes in Jose's life. He still was shoeless about 1.5 hours per day. Yet the frequency with which he wore the new shoes indicates that the shoes donated by TOMS were nevertheless a welcome and appreciated gift."

IV. SUMMARY AND CONCLUSIONS

Our study carries out difference-in-difference estimations on data from 1,578 children in a field experiment in El Salvador that estimates the impact of shoe donations on key outcome variables for these children. Given the widespread use of anecdotes by organizations that interface between donors and the overseas poor, we introduce the concept of a median impact narrative, which may be used by researchers to more effectively communicate study results and by practitioner organizations to convey a truthful account of average treatment effects in the form of narrative.

Given the extraordinary increase in socially conscious firms making donations of in-kind goods in developing countries, we also believe that there are three important practical policy lessons from our TOMS field experiment that are highly relevant to those making in-kind donations:

(1) *Careful targeting and context substantially matter.* This is the key message from our study for socially conscious firms, international donors, and aid agencies. Lower-middle income countries such as El Salvador, where clothes and shoes are relatively widespread and where the government is making significant investments in public health and education, are unlikely to be ideal targets for in-kind gifts in the form of basic consumer items such as shoes and apparel. Targeting must be done carefully so that recipients will clearly benefit from donated items.

(2) *In-kind donations are likely to have unforeseen and unintentional consequences.* In our case, we observe the donated shoes mostly substituting for previously owned shoes, with no decrease in shoelessness among children. Possibly because of the increased time allocated to outdoor activities from the better shoes, we see possibly lower rates of homework and higher rates of bodily injury among children in the treatment group. These injuries may be the natural result of more healthy outdoor activity and may seldom be serious, but we nevertheless find these impacts.

(3) *In-kind donations may exhibit negative externalities on the psychology of recipients, unintentionally fostering a sense of dependency on outside donors.* This is another unintended result we find in our data, and because we only find evidence of it from attitudinal questions rather than observed behaviors, it is a phenomenon that we would encourage other researchers to investigate in impact studies of in-kind goods in order to ascertain its external validity. In the case of donated goods that realize substantial positive impacts on beneficiaries, the question of whether any negative effects of aid dependency outweigh the positive benefits is a difficult one to assess as it requires the difficult comparison of material or physical benefits with psychological impacts.

Our impact study of the TOMS giving program is highly contextual and does not imply that shoe or other types of in-kind donations are unimportant to children universally. Because context matters, shoe donations may realize positive and significant impacts in countries where shoelessness is a genuine barrier to school attendance and/or where the prevalence of foot-borne diseases and parasites such as helminths (hookworm) is high. In other words, the results of this study should not necessarily

be taken as externally valid to countries in sub-Saharan Africa, for example, where hook worm is more common, and there might be very tangible health benefits to providing children with donated shoes.

Nevertheless, millions of shoe donations today are carried out in countries quite similar to El Salvador, where shoe ownership before the intervention is not as high as in developed countries, but still widespread. The astounding growth of socially oriented companies that seek to have an impact on the poor in developing countries through in-kind donations must understand that, absent careful study, in-kind donations are likely to meet with negligible impacts on purported beneficiaries. This causes us to emphasize the importance of careful ground research in a potential recipient area before in-kind donations are made so that donations can be targeted specifically and appropriately. We also emphasize the use of donations as incentives for positive behavior in areas such as school attendance, obtaining vaccinations, health check-ups, and achieving different types of goals. Distributing shoes and other donated goods as rewards for positive behavior may reduce feelings of entitlement and promote a sense of accomplishment, likely mitigating some of the increases in external dependency we find in a context where the shoes were distributed uniformly. Experimental research comparing ad hoc distribution to one in which donations were made in the context of incentives would be a valuable contribution to the debate on the efficacy of international in-kind donations.

REFERENCES

- Afridi, F. 2011. "The Impact of School Meals on School Participation: Evidence from Rural India." *Journal Of Development Studies*, 47 (11): 1636–56.
- Ahmed, A. U., A. R. Quisumbing, M. Nasreen, J. F. Hoddinott, and E. Bryan. 2009. "Comparing Food and Cash Transfers to the Ultra Poor in Bangladesh." International Food Policy Research Institute (IFPRI) Research Monograph 163.
- 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: 1481–95.
- Angrist, J., and J. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Ayode, D., C. M. McBride, H. D. de Heer, E. Watanabe, T. Gebreyesus, A. Tora, G. Tadele, and G. Davey. 2013. "A Qualitative Study Exploring Barriers Related to Use of Footwear in Rural Highland Ethiopia: Implications for Neglected Tropical Disease Control." *PLOS Neglected Tropical Diseases* 7 (4): 1–8.
- Bal, M. and M. Veltkamp. 2013. "How Does Fiction Reading Influence Empathy? An Experimental Investigation on the Role of Emotional Transportation." *PLoS ONE* 8 (1): e55341. doi:10.1371.
- Blattman, C., and P. Niehaus. 2014. "Show Them the Money: Why Giving Cash Helps Alleviate Poverty." *Foreign Affairs*, May/June.
- Blattman, C., N. Fiala, and S. Martinez. 2012. "Employment Generation in Rural Africa: Mid-Term Results from an Experimental Evaluation of the Youth Opportunities Program in Northern Uganda." DIW Berlin Discussion Paper 1201.
- Barrett, C. 2006. "Food Aid's Intended and Unintended Consequences." ESA Working Paper No. 06–05. Rome: Agricultural and Development Economics Division, Food and Agriculture Organization of the United Nations.
- Burwen, J. and D. I. Levine. 2012. "A Rapid Assessment Randomized-Controlled Trial of Improved Cookstoves in Rural Ghana." *Energy for Sustainable Development*, 16 (3): 328–38.
- Colin C., D. Miller, and J. Gelbach. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 90 (3): 414–27.
- Casey, K., R. Glennerster, and E. Miguel. 2012. "Reshaping Institutions: Evidence on Impacts Using a Pre-Analysis Plan." *Quarterly Journal Of Economics* 127 (4): 1755–812.
- Cohen, J., and P. Dupas. 2010. "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment." *Quarterly Journal of Economics* 125 (1): 1–45.
- Cuna, J. 2014. "Testing Paternalism: Cash versus In-Kind Transfers." *American Economic Journal: Applied Economics* 6 (2): 195–230.
- Currie, J. and F. Gahvari. 2008. "Transfers in Cash and In-Kind: Theory Meets the Data." *Journal of Economic Literature* 46 (2): 333–38.
- Cunha, J., G. De Giorgi, and S. Jayachandran. 2011. "The Price Effects of Cash Versus In-Kind Transfers." National Bureau of Economic Research Working Paper 17456.
- Cristia, J., P. Ibrarran, S. Cueto, A. Santiago, and E. Severin. 2012. "Technology and Child Development: Evidence from the One Laptop per Child Program." IZA working paper 6104, Bonn, Germany.

- Currie, J. and F. Gahvari. 2008. "Transfers in Cash and In-Kind: Theory Meets the Data." *Journal of Economic Literature* 46 (2): 333–83.
- Darolia, R., and B. Wydick. 2011. "The Economics of Parenting, Self-Esteem, and Academic Performance: Theory and a Test." *Economica* 78 (310): 215–39.
- Devoto, F., E. Duflo, P. Dupas, W. Parienté, and V. Pons. 2012. "Happiness on Tap: Piped Water Adoption in Urban Morocco." *American Economic Journal: Economic Policy* 4 (4): 68–99.
- DIGESTYC. 2013. Encuesta de Hogares de Propósitos Múltiples (2012). Dirección General de Estadística y Censos, El Salvador.
- Drèze, J. and G. Kingdon. 2001. "School Participation in Rural India." *Review of Development Economics* 5 (1): 1–24.
- Duflo, E., P. Dupas, and M. Kremer. 2015. "Education, HIV, and Early Fertility: Experimental Evidence from Kenya." *American Economic Review* 105 (9): 2757–97.
- Evans, D., M. Kremer, and M. Ngatia. 2009. "The Impact of Distributing School Uniforms on Children's Education in Kenya." Harvard University Working Paper.
- Evans, D and A. Popova. 2014. "Cash Transfers and Temptation Goods: A Review of Global Evidence." World Bank Policy Research Working Paper, WPS 6886.
- Frazer, G. 2008. "Used-Clothing Donations and Apparel Production in Africa." *Economic Journal* 118 (October): 1764–84.
- Frison, L., and S. Pocock. 1992. "Repeated Measures In Clinical Trials Analysis Using Mean Summary Statistics and Its Implications For Design." *Statistics in Medicine* 11: 1685–704.
- Gentilini, U. 2014. *Our Daily Bread*: "What is The Evidence on Comparing Cash Versus Food Transfers?" Social protection and labor discussion paper 1420. Washington, DC: World Bank Group.
- Gertler, P., S. Martinez, P. Premand, L. Rawlings, and C. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington, DC: The World Bank.
- Grider, J. and B. Wydick. 2016. Wheels of Fortune: The Impact of Wheelchair Provision in Ethiopia." *Journal of Development Effectiveness* 8 (1): 44–66.
- Gundersen, C., B. Kreider, and J. Pepper. 2012. The Impact of the National School Lunch Program on Child Health: A Nonparametric Bounds Analysis. *Journal Of Econometrics* 166 (1): 79–91.
- Haushofer, J., and J. Shapiro. "Impacts of Unconditional Cash Transfers". *Quarterly Journal of Economics* (forthcoming)
- Hsee, C., Y. Yang, X. Zheng, and Hanwei Wang. -2015-. "Lay Rationalism: Individual Differences in Using Reason Versus Feelings to Guide Decisions." *Journal of Marketing Research* 52 (1): 134–46.
- Hidalgo, D., M. Onofa, H. Oosterbeek, and J. Ponce. 2013. "Can Provision of Free School Uniforms Harm Attendance? Evidence from Ecuador." *Journal of Development Economics* 103: 43–51.
- Hoffman, V., C. Barrett, and D. Just. 2009. "Do Free Goods Stick to Poor Households? Experimental Evidence on Insecticide Treated Bednets." *World Development* 37 (3): 607–17.
- Kling, J. R., J. B. Liebman, and L. F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.
- Kremer, M. and C. Vermeersch. 2004. "School Meals, Educational Attainment, and School Competition: Evidence from a Randomized Evaluation." World Bank Policy Research Paper 3523.

- Ludwinski, D., K. Moriarty, and B. Wydick. 2011. "Environmental and Health Impacts From the Introduction of Improved Wood Stoves: Evidence from a Field Experiment in Guatemala." *Environment, Development, and Sustainability* 13: 657–76.
- McKenzie, D. 2012. "Beyond Baseline and Follow-up: The Case for More T in Experiments." *Journal of Development Economics* 99: 210–21.
- Mills, F. 2012. "Education Reform Gets High Marks in El Salvador." Council on Hemispheric Affairs Policy Memo #1.
- Rawlins, R., S. Pimkina, C. Barrett, S. Pedersen, and B. Wydick. 2014. "Got Milk? The Impact Of Heifer International's Livestock Donation Programs In Rwanda on Nutritional Outcomes." *Food Policy* 44 (2): 202–13.
- Small, D., G. Loewenstein, and P. Slovic. 2007. "Sympathy and Callousness: The Impact Of Deliberative Thought on Donations to Identifiable and Statistical Victims." *Organizational Behavior and Human Decision Processes* 102 (2): 143–53.
- Vermeersch, C., and M. Kremer. 2005. "School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation." World Bank Policy Research Working Paper Series 2523.
- Wu, C.F.J. 1986. "Jackknife, Bootstrap, and Other Resampling Methods in Regression Analysis." *Annals of Statistics* 14: 1261–350.
- Wydick, B. 2015. "Impact as Narrative." *Development Impact: News, Views, Methods, and Insights from the World of Impact Evaluation*, World Bank Impact Blog, January 7, 2015.
- Wydick, B., B. Janet, and E. Katz. 2014. "Do In-kind Transfers Damage Local Markets? The Case of TOMS Shoe Donations in El Salvador." *Journal of Development Effectiveness* 6 (3): 249–267.

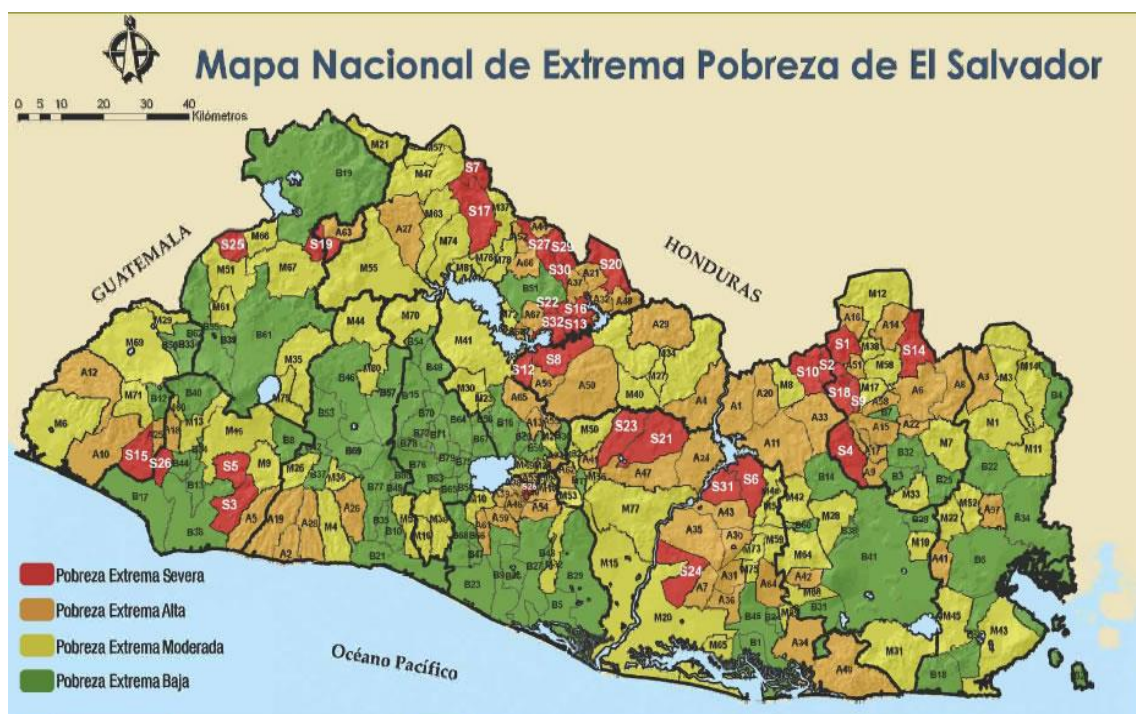
FIGURE TITLES AND SOURCES

Figure 1. TOMS Donation Shoes



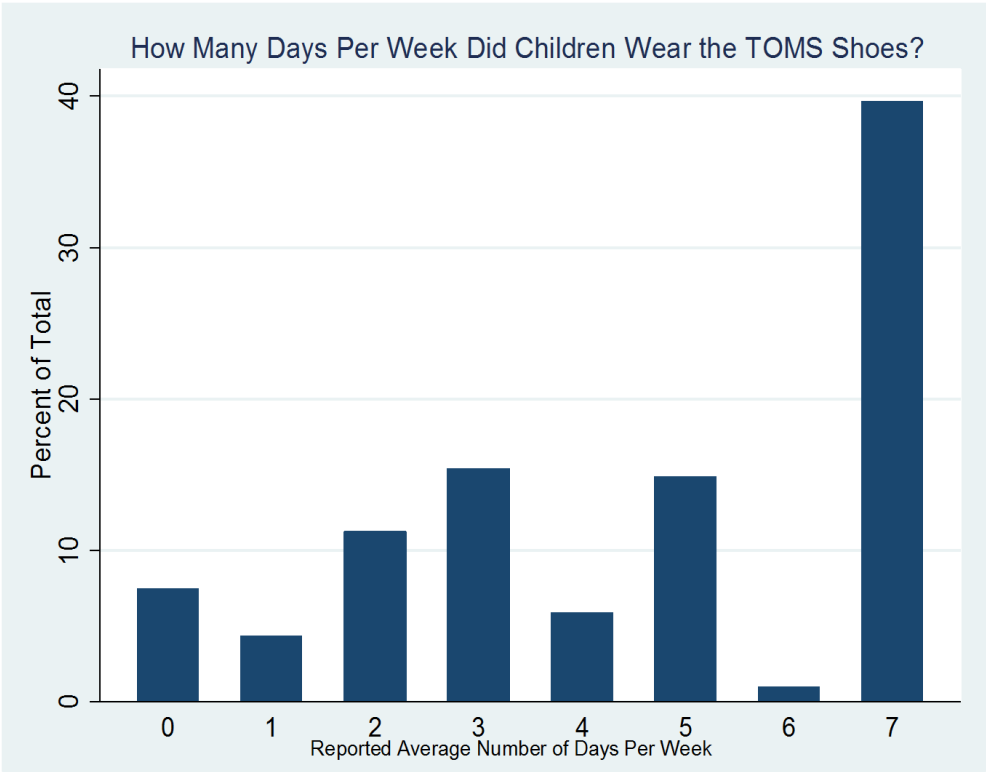
Source: <http://www.toms.com/what-we-give-shoes>

Figure 2. El Salvador Extreme Poverty Map



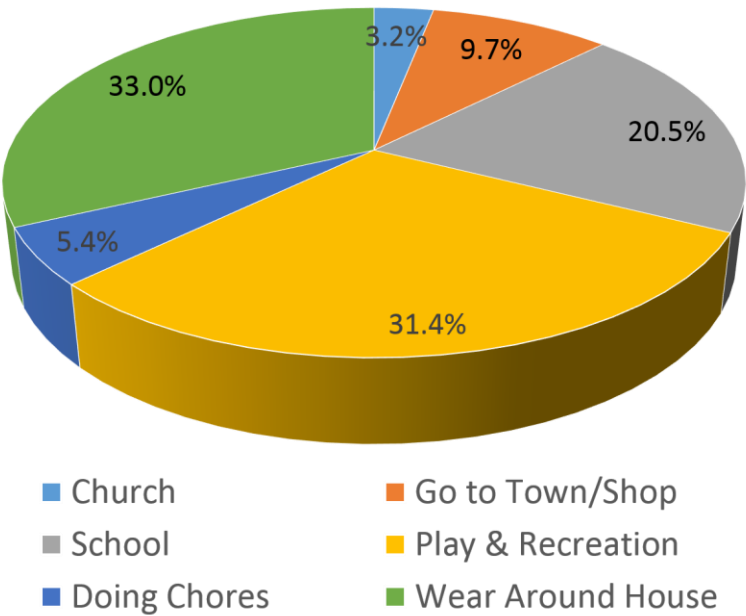
Source: Fondo de Inversión Social para el Desarrollo Local (FISDL)

Figure 3. Days per Week Wearing Donated Shoes



Source: Authors’ analysis based on survey data

Figure 4. Daily Activities While Wearing Donated Shoes



Source: Authors’ analysis based on survey data

Table 1. Covariates between Treated and Control Communities
(Baseline Survey Data)

Variable	Control	Treatment	<i>p</i> -value
	Mean	Mean	
Age of children	9.486	9.332	0.114
Gender of children (Male =1)	0.545	0.497	0.073*
Head of household works in agriculture	0.462	0.522	0.435
Highest level of education, adults in household	5.836	5.346	0.406
Dwelling Index by Household	0.492	0.593	0.447
Consumer durable index by household	0.441	0.467	0.829
Shoe ownership of children	2.060	1.825	0.213
Hours children shoeless during waking hours	2.090	1.963	0.917
Percent shoeless (children owning no shoes)	0.0074	0.0076	0.978
Missed school days by child	0.701	0.886	0.465
Time sleeping	10.68	9.98	0.388
Time eating	1.931	1.934	0.986
Time washing	0.732	0.835	0.228
Time in school	4.613	4.154	0.168
Time working	0.409	0.572	0.082*
Time shopping	0.199	0.348	0.089*
Time doing household chores	0.827	0.749	0.577
Time fetching water	0.144	0.259	0.094*
Time collecting firewood	0.206	0.289	0.367

Time doing homework	0.951	1.110	0.123
Time playing outdoors	1.779	1.733	0.829
Time watching television	1.272	1.513	0.772
Health index among children	0.137	0.060	0.701
Foot health index among children	-0.259	0.189	0.101
Total observations	666	912	1,578

P-values are from simple t-tests adjusted for intra-cluster correlation for significant differences between control and treatment means at baseline. Note: Time allocation data does not control for seasonality at month of survey, where seasonal effects of time allocation are controlled for in regressions.

Source: Authors' analysis based on survey data

Table 2. Impact of TOMS Intervention on Shoe Ownership and Shoelessness

	(1)	(2)	(3)	(4)
	Shoe ownership		Daily hours shoeless	
Variables	Difference-in- differences	ANCOVA	Difference-in- differences	ANCOVA
Treated (received pair of TOMS Shoes)	0.219 (0.181) [0.197]	0.075 (0.061) [0.065]	0.399 (0.717) [0.831]	0.643 (0.447) [0.545]
Observations	2,949	1,302	1,546	541
R-squared	0.115	0.199	0.029	0.134
Baseline control mean:	2.06		2.09	
Baseline control standard deviation:	1.02		2.66	
<p>Difference-in-difference estimations include controls for Ever Treated, Round 2, Age, Sex, Occupation of Household Head, Education of Household Head, Dwelling Quality Index, and month dummies to account for seasonality of schooling and other activities. ANCOVA estimations control for baseline outcomes and same set of control variables. Standard errors clustered at the community level in parentheses. Wild-bootstrapped standard errors clustered at the community level in brackets. Regressions include fixed effects at the Area Development Program (ADP) level. ****p < 0.01, **p < 0.05, *p < 0.10</p> <p><i>Source:</i> Authors' analysis based on survey data</p>				

Table 3. Time Allocation (Seemingly Unrelated Regressions)

(Units in hours)

	(1)	(2)	(3)	(4)	(5)	(6)	
	Working						
Estimation:	Sleeping	Eating	Washing	School	g	Shopping	
Differences-in-differences:	0.436	0.113	0.052	0.472	0.038	-0.211	
(Treated x Round 2)	(0.775)	(0.157)	(0.139)	(0.470)	(0.109)	(0.117)	
Observations	1,562	1,562	1,562	1,562	1,562	1,562	
R-squared	0.18	0.16	0.071	0.66	0.085	0.051	
ANCOVA:	-0.156	0.150	0.059	-0.175	0.348	-0.108	
(Treated)	(0.249)	(0.134)	(0.159)	(0.638)	(0.243)	(0.097)	
Observations	556	556	556	556	556	556	
R-squared	0.093	0.17	0.11	0.68	0.14	0.091	
Baseline control mean hours:	10.68	1.931	0.732	4.613	0.409	0.199	
Baseline control std dev:	1.02	0.65	0.55	1.13	0.88	0.65	
	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Estimation	Chores	Homework	Firewood	Water	Playing	TV	Outdoors
Differences-in-Differences:	-0.064	-0.283**	0.028	0.017	0.084	-0.439	-0.135
(Treated x Round 2)	(0.159)	(0.144)	(0.079)	(0.074)	(0.317)	(0.874)	(0.397)
Observations	1,562	1,562	1,562	1,562	1,562	1,562	1,561
R-squared	0.21	0.22	0.087	0.11	0.25	0.18	0.42

ANCOVA:	-0.226	-0.057	-0.013	0.050	-0.014	-0.001	0.337
(Treated)	(0.134)	(0.177)	(0.074)	(0.146)	(0.323)	(0.254)	(0.340)
Observations	556	556	556	556	556	556	556
R-squared	0.23	0.20	0.088	0.18	0.25	0.17	0.41
Baseline control mean hours:	0.827	0.951	0.206	0.144	1.779	1.272	4.595
Baseline control std dev:	1.07	0.71	0.45	0.37	1.28	1.22	1.65

Difference-in-difference estimations include controls for Ever Treated, Round 2, Age, Sex, Occupation of Household Head, Education of Household Head, Dwelling Quality Index, and month dummies to account for seasonality of schooling and other activities. ANCOVA estimations control for baseline outcomes and same set of control variables. Standard errors clustered at the community level in parentheses. Regressions include fixed effects at Area Development Program (ADP) level. ***p < 0.01, **p < 0.05, *p < 0.10

Source: Authors' analysis based on survey data

Table 4. School Attendance (Days of School Missed during Month, OLS)

	(1)	(2)	(3)
	School days	School days	School days
	missed	missed	missed
Estimation:	All	Boys	Girls
Difference-in-differences	-0.290	-0.316	-0.280
(Treated \times Round 2)	(0.327)	(0.357)	(0.308)
	[0.267]	[0.391]	[0.363]
Observations	2,173	1,152	1,021
R-squared	0.045	0.051	0.039
ANCOVA	-0.165**	-0.195	-0.120
(Treated)	(0.057)	(0.086)	(0.072)
	[0.079]	[0.124]	[0.095]
Observations	664	370	294
R-squared	0.186	0.162	0.238
Baseline control mean:	0.701	0.805	0.595
Baseline control std dev:	1.31	1.33	1.27
<p>Difference-in-difference estimations include controls for Ever Treated, Round 2, Age, Sex, Occupation of Household Head, Education of Household Head, Dwelling Quality Index, and seasonality dummy. ANCOVA estimations control for baseline outcomes and same set of control variables. Standard errors clustered at the community level in parentheses.</p> <p>Wild-bootstrapped standard errors clustered at the community level in brackets.</p> <p>Regressions include fixed effects at the Area Development Program (ADP) level. *p < 0.10, **p < 0.05, ***p < 0.01.</p>			

Source: Authors' analysis based on survey data

Table 5. Foot Health Regressions (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Infectio		Missing			Foot
Estimation:	Cut	n	Irritation	Toenail	Blister	Sores	health
							index
Diff-in-diff:	0.014	0.017	0.043	0.009	0.044	0.033	-0.219
(Treated \times Round 2)	(0.062)	(0.027)	(0.041)	(0.010)	(0.033)	(0.032)	(0.205)
	[0.064				[0.037	[0.039	[0.221]
]	[0.030]	[0.042]	[0.010]]]	
Observations	2,817	2,815	2,817	2,816	2,817	2,814	3,057
R-squared	0.025	0.015	0.004	0.007	0.010	0.011	0.015
ANCOVA:	-0.015	-0.009	0.011	0.001	0.018	-0.011	0.000
(Treated)	(0.051)	(0.010)	(0.027)	(0.005)	(0.024)	(0.029)	(0.146)
	[0.079				[0.027	[0.035	
]	[0.013]	[0.039]	[0.011]]]	[0.200]
Observations	1,272	1,270	1,272	1,272	1,272	1,269	1,406
R-squared	0.153	0.015	0.046	0.018	0.041	0.019	0.081
Baseline control mean:	0.115	0.062	0.088	0.023	0.077	0.072	-0.141
Baseline control std. dev:	0.31	0.24	0.28	0.15	0.27	0.26	1.23

Difference-in-difference estimations include controls for Ever Treated, Round 2, Age, Sex, Occupation of Household Head, Education of Household Head, Dwelling Quality Index, and seasonality dummy. ANCOVA estimations control for baseline outcomes and same set of control variables. Standard errors clustered at the community level in parentheses. Wild-bootstrapped standard errors clustered at the community level in brackets. Regressions include fixed effects at the Area Development Program (ADP) level. * $p < 0.10$, ** $p < 0.05$,

*** $p < 0.01$.

Source: Authors' analysis based on survey data

Table 6. Health Regressions (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Abdominal								
Variables	Headache	Pain	Fever	Dizziness	Injury	Diarrhea	Flu	Skin	Health Index
Diff-in-Differences	-0.001	0.013	0.030	-0.040	0.027	0.020	-0.020	0.011	0.026
(Treated \times Round 2)	(0.006)	(0.061)	(0.061)	(0.036)	(0.024)	(0.026)	(0.049)	(0.046)	(0.193)
	[0.005]	[0.054]	[0.071]	[0.047]	[0.026]	[0.033]	[0.054]	[0.046]	[0.371]
Observations	2,845	2,842	2,844	2,841	2,844	2,845	2,846	2,841	3,069
R-squared	0.024	0.007	0.016	0.014	0.005	0.008	0.010	0.005	0.007
ANCOVA	0.004	-0.005	0.028	-0.044**	0.019	0.036	-0.005	0.017	0.031
(Treated)	(0.051)	(0.033)	(0.025)	(0.016)	(0.014)	(0.019)	(0.037)	(0.028)	(0.096)
	[0.027]	[0.045]	[0.029]	[0.021]	[0.016]	[0.022]	[0.038]	[0.033]	[0.107]
Observations	1,286	1,283	1,285	1,282	1,285	1,286	1,287	1,282	1,406
R-squared	0.083	0.060	0.041	0.041	0.030	0.033	0.030	0.039	0.073
BL control mean:	0.462	0.338	0.293	0.099	0.065	0.076	0.663	0.130	0.075
BL control SD:	0.50	0.47	0.46	0.30	0.25	0.27	0.47	0.31	1.04
Difference-in-difference estimations include controls for Ever Treated, Round 2, Age, Sex, Occupation of Household Head, Education of Household Head, Dwelling									

Quality Index, and seasonality dummy. ANCOVA estimations control for baseline outcomes and same set of control variables Standard errors clustered at the community level in parentheses. Wild-bootstrapped standard errors clustered at the community level in brackets. Regressions include fixed effects at the Area Development Program (ADP) level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Source: Authors' analysis based on survey data

Table 7. Self-esteem OLS Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Self-esteem 1 (+)	Self-esteem 2 (+)	Self-esteem 3 (-)	Self-esteem 4 (+)	Self-esteem 5 (-)	Self-esteem Index [†]
(Treated)	0.017	0.018	0.133*	0.027**	-0.008	0.001
	(0.016)	(0.032)	(0.064)	(0.012)	(0.063)	(0.110)
	[0.023]	[0.035]	[0.078]	[0.014]	[0.062]	[0.111]
Observations	878	889	873	897	862	734
R-squared	0.006	0.004	0.026	0.010	0.006	0.007
Control mean:	0.957	0.892	0.418	0.945	0.489	-0.022
Control std. dev.:	0.20	0.31	0.49	0.23	0.50	1.05

Estimations include controls for Ever Treated, Round 2, Age, Sex, Occupation of Household Head, Education of Household Head, Dwelling Quality Index, and seasonality dummy. SE 1: Child feels of equal value to others; SE 2: Child feels capable of accomplishing things; SE 3: Child feels has nothing to be proud of; SE 4: Child feels satisfied with him or herself; SE 5: Child feels not good at anything. Standard errors clustered at the community level in parentheses.

Wild-bootstrapped standard errors clustered at the community level in brackets. [†]Index includes questions on aid dependency; responses re-scaled so that a higher value in index indicates higher self-esteem. Regressions include fixed effects at the Area Development Program (ADP) level. Standard errors clustered at the community level in round parentheses. ***p < 0.01, **p < 0.05, *p < 0.10

Source: Authors' analysis based on survey data

Table 8. Impacts on Aid Dependence



	(1)	(2)	(3)
Variables	Child believes each family should provide for its own needs.	Child believes others should provide for needs of family.	Child believes family should provide for its own needs AND does not believe that others should provide for family's needs.
Treated	-0.045 (0.037) [-0.040]	0.122** (0.053) [0.047]	-0.129** (0.046) [-0.550]
Age	0.000 (0.004)	-0.000 (0.007)	-0.001 (0.007)
Sex	-0.021 (0.013)	-0.026 (0.019)	0.023 (0.020)
Head of household works in agriculture	-0.038* (0.022)	-0.015 (0.036)	0.006 (0.037)
Maximum education head of household	0.005** (0.002)	-0.008 (0.005)	0.009* (0.005)
Anderson Dwelling Index	-0.000 (0.006)	-0.006 (0.017)	0.009 (0.017)
Household Anderson Consumer good Index	0.002 (0.006)	-0.041** (0.019)	0.041** (0.019)
Constant	0.975***	0.722***	0.289***

	(0.052)	(0.072)	(0.072)
Observations	895	874	867
R-squared	0.030	0.038	0.043
Control Mean:	0.979	0.634	0.368
Control Standard Deviation:	0.21	0.48	0.48

Standard errors clustered at the community level in parentheses. Wild-bootstrapped standard errors clustered at the community level in brackets. Regressions include fixed effects at the Area Development Program (ADP) level. ***p < 0.01, **p < 0.05, *p < 0.10

Source: Authors' analysis based on survey data

APPENDIX A: TIME USE DIARY

Formulario de Registro de Tránsito: **Nombre de la Empresa:** _____ **Dirección:** _____ **Logo:**  **UNIVERSITY OF SAN FRANCISCO** 

Nombre de la Ruta: _____ **Nº de Ruta (indicado en el boleto):** _____ **Nº de Cuentas de Ruta:** _____

Instrucciones: El formulario debe ser llenado por una persona que sea responsable de la ruta. Marque en los espacios los correspondientes.

[illegible]

Appendix B: Pre-Analysis Plan, Children's Shoe Impact

Project name: TOMS Shoes Impact Study

Hypothesis Document: Pre-Analysis Plan_Wydict_2-12-13.pdf

Date submitted: February, 12, 2013 at 6:32:03 PM EST

Submitted by: Bruce Wydict

SHA1 checksum: ea362150c88a4f8241aebc79fe8b49c5d16bb75c

MD5 checksum: 97b4e11db723bfa2d7405dd90fbaa02c

Stable URL: <http://www.povertyactionlab.org/Hypothesis-Registry>

Pre-Analysis Plan: TOMS Shoes Impact Study

Principal Investigators:

Elizabeth Katz, Ph.D. Brendan Janet, M.S., Bruce Wydict, Ph.D., University of San Francisco

Felipe Gutierrez, University of Arizona College of Medicine, Banner Health

Fieldwork Location and Dates of Fieldwork:

El Salvador, January 15, 2012 to February 21, 2013

We follow McKensie's (2012) checklist of articles for a pre-analysis plan suggested for randomized controlled trial studies.

1. Description of the sample to be used in the study:

Our sample consists of households who have children sponsored by World Vision International, who are scheduled to be recipients of TOMS shoe donations living in communities near four Area Development Programs in El Salvador. Randomization of the treatment, dispersal of TOMS shoes, was done at the community level and carried out after the baseline survey. Follow-up survey was undertaken 3 to 4 months after the baseline survey. Households surveyed are a random sample of households in each of these communities, all low-income households with children sponsored by World Vision International. The four ADP regions are chosen to achieve broad coverage of the country of El Salvador geographically. Unit of analysis is at the household level for part A of our market impact study, and at the level of the individual household member in part B of our market impact study. It is at the level of the child (age 6-12) in the life impact study.

2. Key data sources:

Data for the study will come from baseline and follow-up household surveys carried out by field coordinator Brendan Janet and hired enumerators from April 2012 to February 2013. Four Area Development Program regions were surveyed, each containing 4 to 6 village communities. Baseline data were obtained before the experimental intervention, and then follow-up data were taken 3 to 4 months after the intervention. Household heads were interviewed to obtain the data. Data include time diaries of mothers, who record children's activities by hour of the previous day along with background information on every individual in the household. Data also include shoe purchases and the results of the coupon experiment in which we allocated coupons at randomly chosen discounts to treated and

untreated communities to test if redemption is higher in communities that had not received shoe donations.

3. Hypotheses to be tested throughout the causal chain:

- i) H_0/H_a : No impact (positive impact) of receiving donated shoes on child school attendance.
- ii) H_0/H_a : No impact (positive impact) of receiving donated shoes on children's allocation of time toward activities that are facilitated by shoe-wearing. For example, our alternative hypotheses would suggest that children would allocate time away from activities such as TV watching and toward playing sports.
- iii) H_0/H_a : No impact (positive impact) of receiving donated shoes on children's foot health.
- iv) H_0/H_a : No impact (positive impact) of receiving donated shoes on children's self-esteem and psychology.
- v) H_0/H_a : No impact (negative impact) of receiving donated shoes on children's sense that families should provide for their own needs rather than having others provide for them.
- vi) H_0/H_a : No impact (negative impact) of donated shoes on purchases of shoes in local shoe market.

4. How variables will be constructed:

Variables for (i) will be taken from two sources: a) self-reports of missed school days over the last week; and (b) from official school attendance records of school over the previous month, or the last month of school if survey is done over the holiday break.

Variables for (ii) will be taken from mothers' time-diaries about children's activities the previous day. These are obtained from a matrix with activities as rows and hours during the day as columns, where total time during an activity can be summed up over the number of hours with an 'x' in a square. We allow for 2 activities at one time (e.g. play and chores) in which case time is divided in half during the hour over the two activities.

Variables for (iii) will be taken from inspections of children's feet.

Variables for (iv) will be taken from standard self-esteem questions used by child psychologists in our survey, which are given by a Likert scale.

Variables for (v) will be taken from the question on our survey that asks the degree to which children agree with the statement that families should provide for themselves or whether it is the obligation of others to help their family.

Variables for (vi) will be taken from coupon redemption data, from shoe vendors who participated in our experiment. Coupons were either redeemed or not redeemed, and this is our measure of market purchases for children's shoes for part (a) of the market impact study. For part (b) we will compare the difference between children (6-12) purchases during the 3 months between baseline and follow-up with purchases outside this age group, and then compare this difference between treated and untreated communities.

5. Specify the treatment effect equation to be estimated:

For (i) through (iii) we will estimate the following equation that uses difference-in-differences with ADP (region)-level fixed effects:

$$y_{ijt} = \alpha + \mathbf{X}_{ij}'\boldsymbol{\beta} + \theta T + \tau TF + \pi F + \mu_j + \varepsilon_{it}$$

where $\mathbf{X}_{ij}'\boldsymbol{\beta}$ are control variables that describe the child and her household characteristics, which will include age, gender, economic activity of parents, and indices of dwelling quality and asset ownership. T is an indicator of whether the child lives in a treatment community, F denotes an observation in the follow-up period (as opposed to the baseline period), μ_j is an ADP (region)-level fixed effect (which contains 4-6 communities), and ε_{it} is the error term. Impact is captured by the coefficient on the interaction term, τ .

For the impact of children's time from the shoes, we will carry out SUR estimations on time allocated between sleeping, eating, washing and dressing, school, outside work, shopping, housework, collecting water, collecting wood, doing homework, playing, going to church, and watching television. Some of these categories may be combined. We will also examine the health outcomes in questions 15-17, particularly the six foot health impacts in question 17: Cuts, infections, irritations, missing toe nail, blisters, post-blister sore.

For (iv and v), we will use a propensity-score matching to measure differences in children's psychology from questions 32-46 on the survey, since we do not have baseline data. This will include

- Self-esteem questions: Do you feel you are a person of value? Do you feel you are capable of completing things as well as others? Do you feel like there is not much to be proud of? Do you feel satisfied with yourself? Do you sometimes feel like you are not good at anything? Do you believe that each family should provide for their own necessities? Do you believe it's important for others to provide for the necessities of your family?
- Future aspirations questions: Do you feel the future holds good things for you? Do you feel your adult life will be better than that of your parents?

We will create summary indices of these variables to test the hypotheses that the shoe donation program has an effect on families of variables within the area of psychology as well, specifically grouping questions 32-38, 39-40, 41-42-44-46. In addition, on these outcomes we will estimate the simple difference estimation:

$$y_{ij} = \alpha + \mathbf{X}_{ij}'\boldsymbol{\beta} + \theta T + \mu_j + \varepsilon_{it}$$

For (vi) part (A) that uses our coupon experiment, we will estimate

$$y_{hjt} = \alpha + \mathbf{X}_{ij}'\boldsymbol{\beta} + \theta T + \tau TF + \pi F + \mu_j + \varepsilon_{ht}$$

where y is coupon redemption for low-priced/high priced shoes and we index our observations by household, h , instead of i because observations on coupon redemption are at the household level. For part (B) that compares purchases across family members and treated/control communities, we will estimate the diff-in-diff equation:

$$y_{ij} = \alpha + \mathbf{X}_{ij}'\boldsymbol{\beta} + \theta T + \rho TC + \omega C + \mu_j + \varepsilon_i$$

where y_{ij} are shoe purchases during the 3-4 month period before follow-up, T represents being in a treated community and C represents being a member of the children's group (age 6-12) that is a target of the shoe donation. The impact of the donation we would measure then by the coefficient ρ .

6. Plan for how to deal with multiple outcomes and multiple hypothesis testing:

We have several instances in which we have a family of outcomes that can be tested individually and jointly. When testing individually, we will control the family-wise error rate using the Holm-Bonferroni Step-Down procedure. When testing jointly, we will use summary indices over all of the variables in our survey of the same family created in the manner of Casey et al. (2012) and Anderson (2008).

7. Procedures to be used for addressing survey attrition and missing data:

We expect low attrition in the survey, but there will be some attrition in the data due to about 2% of the households refusing to take the survey (because they had already received the shoe donation and had little material incentive to participate in the study further). We will drop these households from the analysis. We will do our best to correct for any missing data at the survey level through follow-up. If we have significant missing variables, we will replace these values with a zero and use a missing variable indicator or drop the control variable. Since we are using a difference-in-difference, we do not expect unchanging state variables (age, gender, etc.) to have strong significance in the estimation, since we are estimating changes over time based on the experimental intervention (shoe donation).

8. Outcomes with limited variation:

We will include some outcome variables with limited variation—for example, in our first ADP, we had very low redemption rates for our coupons, but we will include this ADP in the final data analysis along with the other ADPs in which we had higher redemption rates. We will drop any control variable for which more than 97% of observations carry the same value, including dummy variables.