



Annual Review of Economics

Using Randomized Controlled Trials to Estimate Long-Run Impacts in Development Economics

Adrien Bouguen,¹ Yue Huang,¹ Michael Kremer,^{2,3}
and Edward Miguel^{3,4}

¹Department of Agricultural and Resource Economics, University of California, Berkeley, California 94720, USA; email: abouguen@berkeley.edu, yuehuang@berkeley.edu

²Department of Economics, Harvard University, Cambridge, Massachusetts 02138, USA; email: mkremer@fas.harvard.edu

³National Bureau of Economic Research, Cambridge, Massachusetts 02138, USA

⁴Department of Economics, University of California, Berkeley, California 94720, USA; email: emiguel@berkeley.edu

Annu. Rev. Econ. 2019. 11:523–61

The *Annual Review of Economics* is online at
economics.annualreviews.org

<https://doi.org/10.1146/annurev-economics-080218-030333>

Copyright © 2019 by Annual Reviews.
All rights reserved

Keywords

randomized controlled trials, long-run impacts, panel data, longitudinal data, development economics, cash transfers, child health

Abstract

We assess evidence from randomized controlled trials (RCTs) on long-run economic productivity and living standards in poor countries. We first document that several studies estimate large positive long-run impacts, but that relatively few existing RCTs have been evaluated over the long run. We next present evidence from a systematic survey of existing RCTs, with a focus on cash transfer and child health programs, and show that a meaningful subset can realistically be evaluated for long-run effects. We discuss ways to bridge the gap between the burgeoning number of development RCTs and the limited number that have been followed up to date, including through new panel (longitudinal) data; improved participant tracking methods; alternative research designs; and access to administrative, remote sensing, and cell phone data. We conclude that the rise of development economics RCTs since roughly 2000 provides a novel opportunity to generate high-quality evidence on the long-run drivers of living standards.



1. INTRODUCTION

Development economics is an eclectic and methodologically rich field, featuring important contributions that utilize cross-country data; historical information; administrative records; and, increasingly, original survey data to understand the determinants of long-run living standards in poor countries. The roles played by human capital investments, well-functioning credit markets, and cash transfers have figured prominently in debates within the field (Banerjee et al. 2015c, Gennaioli et al. 2012, Haushofer & Shapiro 2016), as have specific programs and policies that combine elements of these and other approaches (Banerjee et al. 2015b). Notably, development economists have also pioneered novel field experimental methods over the past 20 years, carrying out thousands of randomized controlled trials (RCTs), often in close collaboration with governments of low-income countries and nongovernmental organizations.

This review article surveys what we have learned about the determinants of long-run living standards from this growing body of RCTs in development economics and argues that these studies provide an exceptional opportunity to generate high-quality evidence on the impacts of a range of international development interventions on economic productivity and living standards.

As quantitative evidence on the rapid increase in the use of RCTs in development economics over the past 20 years, **Figure 1** presents the cumulative number of registered RCTs in development economics conducted from 1995 to 2015 from the American Economic Association (AEA) RCT Registry.¹ These data are likely to be lower bounds on the total number of relevant studies,

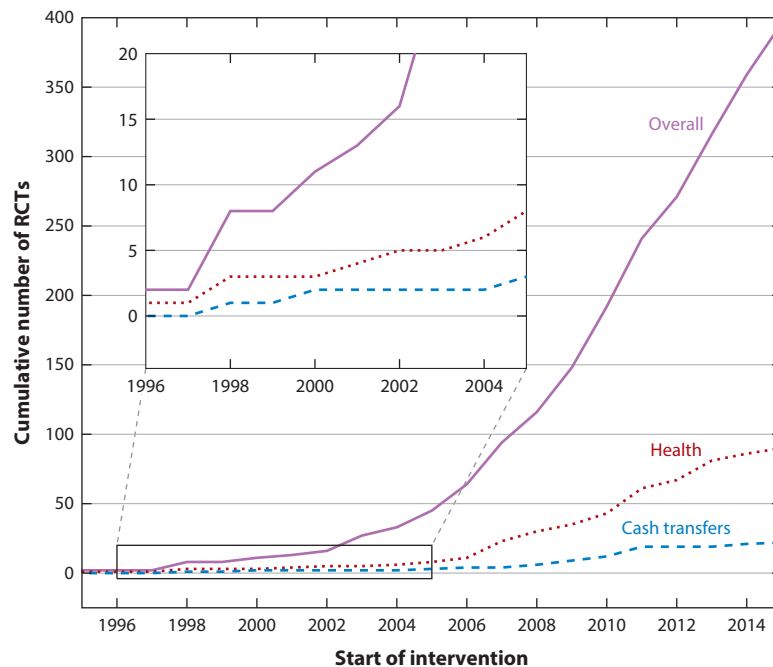


Figure 1

Cumulative number of completed randomized controlled trials (RCTs) in low- and middle-income countries from 1995 to 2015 in the American Economic Association's RCT Registry (<https://www.socialscienceregistry.org>).

¹Data were extracted from the AEA Registry (<https://www.socialscienceregistry.org>) on August 1, 2018. We extracted studies conducted in low- and middle-income countries, according to World Bank definitions.

since not all RCTs are registered, but the patterns in the data remain clear. Following influential early RCTs, such as the Mexico PROGRESA study (Skoufias & McClafferty 2001), the Kenya deworming study (Miguel & Kremer 2004), and the early education studies described by Kremer (2003), there was a surge in the use of RCTs in development economics in the decade of the 2000s. The data indicate that cash transfer programs and health interventions constitute a large share of these studies.² Our impression as researchers active in the field is that the pace of new RCTs in development economics has only accelerated since 2015.

The timing of this surge in development economics RCTs opens up an intriguing possibility: It has been roughly 20 years since these early interventions from the late 1990s and early 2000s were conducted, allowing researchers to begin to assess truly long-run impacts. For child health and education programs, beneficiaries of the early RCTs are now adults, allowing an assessment of long-run impacts on labor productivity, consumption, and living standards. Given the large numbers of RCTs launched in the 2000s, every year that goes by means that more and more RCT studies are aging into a phase where the assessment of long-run impacts becomes possible.

Beyond the opportunity presented by early RCTs in development economics, there are also many RCTs conducted by researchers in international public health that are promising. **Figure 2** presents the cumulative number of RCT studies in the cash transfer and child health areas that have been published in both public health and economics journals, taken from the AidGrade database.³ It is apparent that experimental research methods were widely adopted at least a decade earlier in public health (and related fields) than in economics, and that public health features an even larger body of evidence in terms of the raw count of studies.

Several of these studies generated exogenous variation in child nutrition and health that has already laid the groundwork for long-run evaluations. For instance, the famous INCAP child nutrition experiment in Guatemala was initiated in 1969 by public health researchers (Martorell et al. 1995), and 35 years later, economists followed up on the original sample and estimated significant gains in male wages, improved cognitive skills, and even some positive intergenerational effects (Behrman et al. 2009, Hoddinott et al. 2008, Maluccio et al. 2009), as we detail below.

Some early studies have the advantage of observing truly long-term changes, and have generated invaluable insights into the economic mechanisms underlying program effects, as well as documenting broader technological and institutional changes. In 1974–1975, Christopher Bliss and Nicholas Stern led an extraordinary data collection effort in an Indian village, Palanpur. Along with other researchers, they surveyed this village intensively across several decades and documented how lives and livelihoods changed from the 1970s to today, including local experiences with the Green Revolution and intensification of agriculture, structural transformation, and increasing market integration (Himanshu et al. 2018). A limitation of these early studies is their relatively small sample size of households and focus on a specific geographic region.

We focus on completed studies and do not include RCTs that are ongoing, in the design phase, or withdrawn. This database is not comprehensive, since study registration only became the norm approximately a decade after RCTs became common in development economics, and registration is voluntary; that said, many research institutions and journals are actively promoting registration of both ongoing and completed studies. We omit counts of studies registered after 2015, since most such studies are ongoing.

²We consider a trial to be in the cash transfer or health category if its abstract contains the keywords “cash transfer” or “health,” respectively.

³We extracted data from the AidGrade project (<http://www.aidgrade.org>) in August 2018. AidGrade is a meta-analysis database focusing on 10 types of development aid programs. Notably, they use a somewhat different definition of sector than ours: They restrict attention to health interventions in deworming, HIV/AIDS education, micronutrients, school meals, bed nets, safe water storage, and water treatment, and not all studies focus specifically on children.

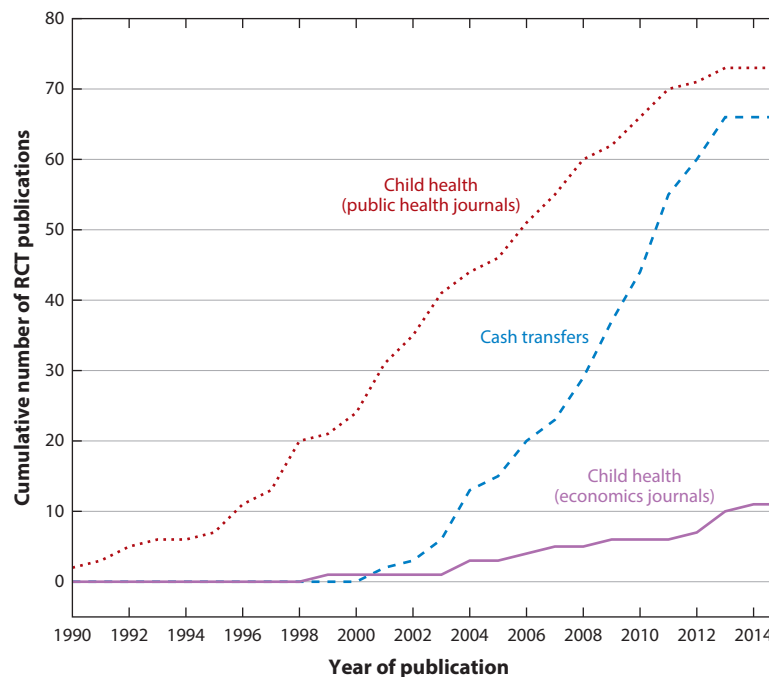


Figure 2

Cumulative number of randomized controlled trial (RCT) publications on cash transfers and child health in the AidGrade database (<http://www.aidgrade.org>).

Another route to assess long-term program or policy impacts that has been common is to exploit natural experiments. This strategy is common in the economic history literature. For instance, Bleakley (2007) exploits the introduction of a hookworm-eradication campaign in the US South, combined with the cross-area differences in pretreatment infection rates, to form an identification strategy and shows that the eradication campaign had long-lasting impacts on income and return to schooling. Similarly, Acemoglu & Johnson (2007) exploit the international epidemiological transition, which led to potentially exogenous differential changes in mortality from tuberculosis, pneumonia, malaria, and various other diseases; countries with larger baseline disease burden thus experience a larger reduction in mortality. Using predicted mortality as an instrument for life expectancy, the authors estimate the effects of life expectancy on population and GDP. Almond (2006) alternatively exploits the timing of the 1918 influenza pandemic, which arrived unexpectedly in the fall of 1918 and had largely subsided by January 1919, and shows large negative educational and labor market effects on cohorts in utero during the pandemic. Evidence from natural experiments is particularly compelling when the policy or program variation studied is truly random. Bleakley & Ferrie (2013), for example, take advantage of the 1832 Cherokee Land Lottery in the state of Georgia to assess the long-term impact of large shocks to wealth. These studies provide valuable insights into the long-term impacts of various cash transfer or health interventions, but natural experiments such as these are hard to come by. Rich historical census or other records are also necessary for researchers to link participants' treatment status to later outcomes, and those records tend to be less available in low-income contexts.

Changes from natural policy variations have also made it possible for researchers to study long-run impacts of cash transfer and child health interventions in wealthy countries. This literature

generally shows large, positive, and persistent effects of such programs; while we briefly discuss it, the extensive literature in high-income countries is not the focus of this article, and we refer interested readers to the recent surveys by Almond et al. (2017) and Hoynes & Schanzenbach (2018). In the United States, the Earned Income Tax Credit (EITC) program has been shown to improve beneficiaries' academic achievement, education attainment, employment, and earnings in the long run (Bastian & Micheltmore 2018, Chetty et al. 2011b). Bastian & Micheltmore (2018) estimate that an additional \$1,000 in EITC exposure when a child is 13–18 years old increases adult earnings by 2.2%, with the primary channel being induced increases in pretax family earnings. Similarly, the Food Stamp Program improved child health in the medium run (East 2019); reduced metabolic syndrome conditions such as obesity, heart disease, and diabetes in adulthood; and increased long-term education and earnings for women (Hoynes et al. 2016). The US Mothers' Pension program, implemented during 1911–1935, benefited the male children of the recipients up to 70 years later: The program increased longevity by one year, reduced the probability of being underweight by half, increased educational attainment by 0.34 years, and increased income in early adulthood by 14% on average, all substantial gains (Aizer et al. 2016). A notable exception is the Seattle-Denver Income Maintenance Experiment, which did not appear to generate long-run benefits (Price & Song 2016). Public health interventions in the United States, especially in early childhood, also generate long-run gains. For instance, the successful hookworm-eradication campaign in the US South increased school enrollment, attendance, literacy, and income roughly 30 years later (Bleakley 2007).⁴ Many of these studies exploit large-scale policy changes and leverage rich administrative data and census records, which are often not readily available in poor countries. As a result, much of the evidence from development economics that we survey in this article relies on original data collection, often including household surveys.

The remainder of this article proceeds as follows.

Section 2 summarizes and evaluates the growing body of evidence from RCTs on the long-term impacts of international development interventions and finds that most (though not all) provide evidence for positive and meaningful effects on individual economic productivity and living standards. Most of these studies examine existing cash transfer, child health, or education interventions and shed light on important theoretical issues such as the existence of poverty traps (Bandiera et al. 2018) and returns to human capital investments in the long term. One notable pattern in the existing body of evidence is the finding that impacts often differ substantially by respondent gender, arguably as a result of the different educational and labor market opportunities facing females and males in most low-income countries. Another is that several existing human capital investment programs, in both health and education, appear to have high rates of return, making them potentially attractive for public policy. We observe some heterogeneity in rates of return for different age cohorts, echoing the literature on the attractiveness of early childhood interventions (Heckman 2006), but we note that interventions targeting children already in school often present high returns. We caution that the studies we summarize may not be representative of all the relevant interventions because projects that attract enough interest and resources for long-term evaluations can be selected on certain traits by both researchers and donors. Many of these characteristics are unobservable and not well understood, thereby potentially generating publication biases.

⁴Early childhood interventions such as the Perry Preschool project lead to increases in high school graduation and college attendance rates and some positive impacts on economic outcomes, criminal behavior, drug use, and marriage for women (Anderson 2008). The Head Start program significantly reduced child mortality rates (Ludwig & Miller 2007) and improved long-term education and health, closing one-third of the gap between children with median- and bottom-quartile family income (Deming 2009).

Section 3 implements a systematic survey exercise and evaluates which existing RCTs are likely to be amenable to long-term follow-up research, with a particular focus on cash transfer and child health programs, which, as we show above, are particularly abundant in the literature. We first consulted existing meta-analysis and survey articles and extracted hundreds of existing experimental studies in these two areas. We then implemented a rigorous screening procedure to identify the studies that could feasibly—and productively—be followed up in future research, after accounting for research design and data challenges, such as a lack of statistical power, high attrition rates or differential attrition across treatment arms, and phase-in designs that dampen cross-arm differences in program exposure. Fortunately, even after screening, we identified dozens of existing RCTs in the cash transfer and child health areas that appear to be attractive candidates for long-term follow-up studies; we typically use a follow-up period of roughly a decade to denote the long run. We view this identification of studies that appear promising for long-run evaluation as a public good for the development economics research community.

Section 4 presents a methodological discussion on promising approaches to estimating long-term impacts, both among existing RCTs and using approaches that can be taken prospectively to make long-run follow-up surveys more successful. We first discuss the assumptions under which it is possible to identify long-run treatment effects using a phase-in research design. We provide lessons from our experience in conducting long-term tracking studies, as well as innovative data approaches. An important methodological question is whether it is worthwhile to conduct follow-up research for RCTs that demonstrated limited short-run impacts, or whether it is safe to assume that any effects fade out in such cases. We discuss evidence from several existing studies that long-run impacts may exist even in the absence of clear-cut short-run effects. There are plausible conceptual reasons for such a pattern: If education and experience are complements in the labor market, then the magnitude of program impacts can grow over time (Brunello & Comi 2004). Policies aimed at improving education can even have negative impacts on beneficiaries' labor market outcomes in the short term, as they may remain in school or in training, or they may experience a longer job-searching period as they search for certain jobs (such as jobs in the public sector or formal sector). In such cases, the absence of long-run evidence could lead to the erroneous conclusion that the benefits of a human capital investment are small. The need for truly long-run labor market data may be particularly important for females, who often have lower labor market attachment during peak child-bearing years, before fully reentering the labor force in mid-life.

Section 5 discusses the implications of this evidence for development economics. We conclude that the rise of development RCTs over the past two decades provides an exciting opportunity for scientific progress by generating credible evidence on the determinants of living standards over the long run. We predict and hope that the trickle of early studies that exploit RCTs to generate long-run evidence will become a flood in the coming years.

2. WHAT HAVE WE LEARNED? A REVIEW OF THE EXPERIMENTAL EVIDENCE

Relatively little is currently known about the long-run impacts of many common interventions in international development. A systematic review by the World Bank (Tanner et al. 2015) focusing on early childhood interventions was able to identify only a single study that reported later employment and labor market outcomes (Gertler et al. 2014). Contemporaneous work by Molina Millán et al. (2019) focusing on conditional cash transfers also concludes that very few studies are able to confidently assess the later employment and labor market impacts of the transfers, as many beneficiaries are still in school and not yet in the labor force. As full-time students usually have

lower earnings, estimates obtained when only a portion of the participants have entered the labor market could understate the true long-run benefits of an intervention or even get the sign wrong, especially if the intervention increases schooling and delays labor market entry. Moreover, as noted in Section 1, if labor market experience and education are complements, then even early estimates obtained when all participants are in the labor force could understate the true long-run benefits of the intervention, for instance, if individual labor productivity grows more rapidly over time for the more educated. This raises the possibility that very-long-run evaluations may be necessary to confidently assess true program impacts and cost effectiveness.

In this section, we assess the evidence from the emerging body of literature that exploits RCTs to estimate long-run impacts of development interventions. One pattern that emerges from the handful of existing studies is that human capital interventions appear to be particularly effective at boosting long-run economic outcomes. For instance, direct investments in child health, such as deworming (Baird et al. 2016a), nutritional supplementation (Hoddinott et al. 2008), and perinatal interventions (Charpak et al. 2016), have all been found to generate meaningful impacts on adult labor productivity. Certain investments in education, including cognitive stimulation in early childhood (Gertler et al. 2014, Kagitcibasi et al. 2009) and scholarship programs (Bettinger et al. 2018), also yield positive returns. Interventions that aim to improve child education, nutrition, and health by leveraging a conditional cash transfer similarly appear to have persistent effects on earnings in some cases (Barham et al. 2017), although not in others: Molina Millán et al. (2018) find no meaningful impacts, possibly because their sample population is still relatively young.

The other set of RCTs that estimate long-run impacts examine unconditional cash transfers and various entrepreneurial grant assistance programs. These programs typically have quite large short-term effects on labor and firm productivity (see, for example, Blattman et al. 2013). Yet most gains appear to fade out after several years (Araujo et al. 2017; Blattman et al. 2018a,b). Similar patterns are sometimes observed in medium-run follow-up studies (Baird et al. 2016b). One exception, which we discuss further below, is provided by multifaceted programs that provide assets to poor households, as well as training and other forms of support (Bandiera et al. 2017, 2018; Banerjee et al. 2016), which appear to have more persistent effects.

Table 1 summarizes all (to the best of our knowledge) of the studies that satisfy our screening criteria. For inclusion, a study had to

1. be in a relevant category (cash transfer or child health interventions),
2. have randomized treatment,
3. report outcomes at least (roughly) 10 years after the intervention started, and
4. report labor market or living standards outcomes.

While we mainly focus on long-run impacts of cash transfers and child health interventions, for completeness, we also briefly discuss other relevant studies in the main text (although some are omitted from **Table 1** due to our inclusion criteria). In this review, we interpret long-run impacts to be persistent effects on the labor market or living standards outcomes of the program beneficiaries over a period of roughly 10 years. We focus on labor market outcomes because they directly reflect individual productivity and largely determine future household living standards in most cases. Many studies document short- to medium-run schooling gains, but these may or may not translate into higher earnings due to institutional or other constraints; thus, it is important to directly assess labor market outcomes. It is notable that many of the studies discussed in **Table 1** are new unpublished working papers (at the time of writing).

Table 1 Existing evidence on long-run impacts of development RCTs

	Study	Country	Intervention	Start of Intervention	Years to Follow-up	Data	Attrition Rate	Long-Run Impacts
1	INCAP (Hoddinott et al. 2008)	Guatemala	Nutrition	1969	35	Survey	40%	Men experienced a 46% increase in average wages.
2	TEEP (Kagitchbasi et al. 2009)	Turkey	Stimulation	1983	22	Survey	49%	More education and more prestigious jobs. Higher consumption but not statistically significantly.
3	Jamaica (Gertler et al. 2014)	Jamaica	Stimulation	1986	22	Survey	19%	Increased earnings by 25% for beneficiaries who were growth-stunted as toddlers.
4	KMC (Charpak et al. 2016)	Colombia	Perinatal	1993	21	Survey	38%	Increased school attendance, labor force participation, and wages.
5	PACES (Bettinger et al. 2018)	Colombia	Scholarship	1994	20	Admin	3% ^a	An 8% increase in earnings. Effects are driven by those who applied to vocational schools.
6	KLPS (Baird et al. 2016a, 2018)	Kenya	Deworming	1998	15	Survey	16%	More education, especially for females. More nonagricultural self-employment. Increased total earnings by 13% and consumption by 23%, with effects concentrated among men.
7	RPS (Barham et al. 2017, 2018)	Nicaragua	CCT	2000	10	Survey	10%	Positive effects on temporary migration and earnings through increasing education and learning for men and improving nutrition and reproductive health for women.
8	PRAF II (Molina Millán et al. 2018)	Honduras	CCT	2000	13	Survey ^b	NA ^c	Improved education and increased probability of migration. Effects on wages or earnings are unclear. Some negative effects on hours worked for women.
9	BDH (Araujo et al. 2017)	Ecuador	UCT	2004	10	Survey	14%	No impacts on learning, small impacts on education, and no impacts on labor market outcomes.
10	YOP (Blattman et al. 2018b)	Uganda	Grant	2007	9	Survey	13% ^d	No impacts on employment, earnings, and consumption.
11	THP (Banerjee et al. 2016)	India	Grant	2007	7	Survey	<1%	Positive effects on consumption, assets, income, food security, financial stability, time spent working, and physical and mental health.
12	TUP (Bandiera et al. 2017)	Bangladesh	Grant	2007	7	Survey	15%	Positive effects on consumption, productive assets, and savings.
13	Ghana Scholarship (Duflo et al. 2018)	Ghana	Scholarship	2008	8	Survey	5%	More tertiary education, higher probability of obtaining more desirable jobs (e.g., jobs in the public sector; jobs with more benefits), and reduced fertility for women.
14	Bangladesh Child Marriage (Buchmann et al. 2018)	Bangladesh	CCT	2008	9	Survey	13%	The conditional incentives delayed marriage and childbearing, increased education, and had insignificant effects on income-generating activities.

Abbreviations: BDH, Bono de Desarrollo Humano; CCT, conditional cash transfer; INCAP, Institute of Nutrition of Central America and Panama; KLPS, Kenya Life Panel Survey; KMC, Kangaroo Mother Care; NA, not applicable; PACES, Programa de Ampliación de Cobertura de la Educación Secundaria; PRAF II, Programa de Asignación Familiar phase II; RCT, randomized controlled trial; RPS, Red de Protección Social; TEEP, Turkish Early Enrichment Project; TUP, Targeting the Ultra-Poor Program; UCT, unconditional cash transfer; YOP: Youth Opportunities Program; THP, Targeting the Hard Core Poor Program.

^aThe authors obtain identification numbers that are valid for 97% of the sample. These identifications are then matched to each administrative data set, with different match rates.

^bFor labor market outcomes, the authors use national household survey data, assigning treatment status to individuals based on their municipality of birth.

^cAttrition rates are challenging to calculate because new samples were drawn and treatment status was assigned to individuals based on their municipality of birth. These samples do not necessarily correspond to the originally sampled households at the time of the intervention.

^dDifferential attrition was observed across treatment and control groups.

2.1. Long-Run Impacts of Cash Transfers

Cash transfer programs can achieve large persistent impacts if (a) the poor have high returns to physical capital, and business grants relax constraints in their ability to borrow, save, and mitigate risk, thereby improving living standards; or (b) the poor have high returns to human capital, and cash transfers or direct education and health interventions promote investments in education and health, thereby improving living standards. Even if cash transfers do not lead to persistent impacts on consumption, the rate of return and welfare impacts of the programs could potentially be large. Suppose, for example, that people are credit constrained and have a high rate of return to a good purchased with the transfer that is not permanent but persists for several years, e.g., a motorcycle that they use as a taxi or a metal roof that allows them to avoid purchasing grass for thatching, but that this good completely depreciates before the final endline measurement. Suppose also that they allocate all of the income generated by the good before it depreciated into immediate consumption, so there were no persistent welfare gains at the time of endline measurement. If the net present value of the temporary consumption gains due to the transfer were sufficiently large relative to the size of the transfer, then the program may have nonetheless been highly beneficial. The total welfare impacts can be recovered if measurements are collected sufficiently frequently, although this is typically not done. In this section, we assess the accumulating evidence on the magnitude and persistence of the effects of cash transfer programs, beginning with unconditional cash transfer programs.

2.1.1. Unconditional cash transfers. Evidence on the long-term impacts of unconditional cash transfers remains scarce, as they were fairly uncommon in the early wave of development RCTs in the 1990s and early 2000s. One exception is the work of Araujo et al. (2017), who study the long-term effects of the Ecuador Bono de Desarrollo Humano (translation: Human Development Voucher) program. As in many development RCTs, the control group began receiving treatment three years after the treatment group, most likely dampening estimated program impacts and leading estimates to be lower bounds on true effects relative to a trial design with a never-treated control group. The authors find that, 10 years after the program, children in the early treatment group did not demonstrate improved learning outcomes in late childhood. Using a regression discontinuity design to exploit the poverty index cutoff, they also show that cash transfers received in late childhood modestly increased the proportion of young women who completed secondary school but did not affect their education and work choices after graduation. Taken together, there are limited detectable long-run impacts of a fairly generous unconditional cash transfer program.

The fade-out of effects from unconditional cash transfers has also been observed in the medium run by Baird et al. (2016b).⁵ In a cash transfer program with both unconditional and conditional transfer arms in Malawi, they find that female unconditional cash transfer recipients show a modest delay in the timing of marriage, fertility, and HIV infection, but that these effects fade out after roughly two years. There are some differences across the unconditional and conditional cash transfer arms; for instance, for girls who had already dropped out of school when the study began, two years of conditional cash transfers do produce meaningful increases in educational attainment and lead them to marry significantly more educated husbands, which may lead to long-run benefits.

2.1.2. Entrepreneurial grants. The long-run effects of entrepreneurial grant programs are similarly mixed, possibly due to the very heterogeneous nature of the interventions and populations studied. Blattman et al. (2018b) study the Youth Opportunities Program (YOP), an

⁵We omit this study from **Table 1** because it only reports medium-run outcomes at a time horizon of less than a decade.

entrepreneurial grant program launched in Uganda in 2008. The program granted hundreds of small groups \$400 per person to kick-start microenterprises. The program increased average earnings by 38% and consumption by 10% after four years (Blattman et al. 2013), but after nine years, the control group had completely caught up with the treatment group in terms of employment, earnings, and consumption. There are lasting effects on assets and occupational choice, suggesting some persistent economic gains. YOP beneficiaries and their children also show little to no health or education gains, except for modest improvements in physical functioning among children of female recipients. These results are consistent with the findings of Blattman et al. (2018a)⁶ that a program in Ethiopia that provided grants of \$300 plus basic business consulting raised incomes by one-third in the first year, but that employment and earnings largely converged across the treatment and control groups within five years.

While both unconditional cash transfer programs and entrepreneurial grant programs appear to initially help the poor accumulate assets, evidence from the limited number of studies at hand is broadly consistent and indicates that these assets are generally gradually run down over time, generating little permanent impact on poverty. One possible explanation is that neither type of program directly ties the transfers to human capital investments. Notable exceptions are programs that tie asset transfers to more intensive training and support, namely, the multifaceted assistance projects targeted at the extremely poor studied by Bandiera et al. (2017) and Banerjee et al. (2016). Banerjee et al. (2016) find that an asset transfer combined with support for 18 months in India generated impacts that persisted and even grew over seven years. Positive effects are found across all categories of outcomes, including consumption, assets, income, food security, financial stability, time spent working, and physical and mental health. Bandiera et al. (2017) find similar evidence for persistent effects of a program transferring livestock and training, valued at \$1,120 in 2007 purchasing power parity, to the ultra poor in Bangladesh. Like Blattman et al. (2018b), they find persistent effects on assets and occupational choice (particularly livestock rearing) seven years after the program, but in contrast to other work, they also find persistent effects on consumption. Some of the estimates for impacts after seven years are likely to be lower bounds, since the control group was subsequently phased into treatment.

Using the same exogenous variation induced by the nongovernmental organization BRAC's Targeting the Ultra-Poor program (Bandiera et al. 2017), Bandiera et al. (2018) present evidence for the existence of poverty traps in this setting, providing a potential explanation for the persistent gains that they find. Conceptually, poverty traps can occur when there are increasing returns to scale to factors that can be accumulated, and when credit markets are imperfect. They find that, in their sample, individuals' assets exhibit a bimodal distribution: After the intervention, those above a certain threshold accumulate assets at a decreasing rate, while individuals below that threshold lose assets at an increasing rate. This is consistent with theoretical predictions from a poverty trap model with multiple equilibria. While the authors provide evidence for a poverty trap in their setting, few other cash transfer programs seem to generate persistent effects. It is, of course, possible that the assistance provided in other programs was simply too small to move recipients over the threshold needed to escape the poverty trap. Yet both the nongovernmental organization Give Directly (Haushofer & Shapiro 2018) and the 19th century Cherokee Land Lottery (Bleakley & Ferrie 2013) provided very large-scale transfers, and the evidence from neither suggests poverty traps. Moreover, empirical wealth distributions are typically unimodal, rather than following the bimodal distribution observed in Bandiera et al.'s (2018) setting, so the existence of poverty traps may be specific to certain contexts.

⁶We omit this study from **Table 1** because it only reports medium-run outcomes at a time horizon of less than a decade.

What explains the differences in impacts between these ultrapoor programs and other enterprise grant and assistance projects? While there is no definitive answer, there are several plausible interpretations, beyond the possibility that this particular setting and population had conditions that led to a poverty trap. Differences in targeting (i.e., the poor versus the extremely poor) could play a role. Moreover, the multifaceted and intensive training in the ultrapoor programs may have induced greater human capital accumulation or addressed behavioral barriers to saving. Further research is needed to provide more definitive answers.

2.1.3. Conditional cash transfers. Despite the proliferation of evaluations of conditional cash transfer programs, especially in Latin America, high-quality experimental evidence on their long-term impacts remains limited. The estimation of long-run impacts is also complicated by the fact that many RCTs employ phase-in designs, in which the control group later receives treatment, sometimes after only a year or so. As we discuss below (Section 4), phase-in designs are likely to yield lower bounds on the treatment effects that would be obtained with a pure control group (that never received treatment), somewhat changing the interpretation of effect estimates and also making null results harder to interpret.

Barham et al. (2017, 2018) evaluate a three-year conditional cash transfer program in Nicaragua, which was later phased in. They exploit the fact that, although both the early treatment and late treatment groups received three years' worth of cash transfers, in the latter group, the boys largely missed the transfers that were most likely to prevent them from dropping out of school, and the girls missed the transfers during their potentially critical early teenage years, around the onset of puberty. After 10 years, among children aged 9–12 years at the start of the program, young men in the early treatment group showed increased schooling and learning, which translated into more engagement in wage work, higher rates of temporary migration for better paying jobs, and higher earnings; young women in the early treatment group reached sexual maturity later, had lower body mass index scores, and started sexual activities later, resulting in lower overall fertility in young adulthood. Despite modest effects on education and learning outcomes for young women, they experienced similar earnings and labor market participation gains as men.

Buchmann et al. (2018) evaluate a program aimed at reducing child marriage and teenage childbearing and increasing girls' education in Bangladesh. The authors cross-randomized a six-month empowerment program and a financial incentive to delay marriage. Nine years after the program started (and 4.5 years after the end of the program), girls randomized to receive conditional cash incentives got married later and were less likely to bear a child as teenagers. Unlike cash transfer programs that are conditional on school enrolment or attendance, this program also benefits vulnerable girls that are already out of school at baseline. The authors find that empowerment programs alone have no discernible effect on marriage but do improve education. The empowerment program also increases income-generating activities, particularly labor market participation, among older girls.

Evidence from other related interventions is less conclusive. PROGRESA/Oportunidades is the pioneering Mexican conditional cash transfer program that has served as a model for many other programs in Latin America and beyond, and it was evaluated with an RCT in its pilot phase. The program started in 1997 and offered monetary transfers to households conditional on investing in the education, health, and nutrition of the children (e.g., attendance at school, regular clinic visits). Many studies have exploited the experimental design to assess medium- and long-run impacts (for a survey, see Parker & Todd 2017), but the long-run studies are limited by the original evaluation sample's research design: Policy makers decided that the original control group villages would be phased into treatment a mere 18 months after the treatment communities, creating a relatively short gap between the early and late treatment groups. Moreover, participant



attrition in follow-up survey rounds was relatively high and differential across treatment arms. Exploiting this data, Behrman et al. (2011)⁷ show that, in the medium run by 2003 (six years after the start of the program), the greater exposure (of 18 months) to the program in early treatment PROGRESA/Oportunidades communities significantly increases schooling for both genders and decreases participation in the labor force for boys but not for girls.

Given the limitations of the research design and follow-up survey data, long-run impacts of PROGRESA/Oportunidades on labor force participation, wages, and earnings are mostly estimated nonexperimentally⁸ and are mostly large and positive. Adhvaryu et al. (2018) focus on a small cohort who were 18 at the time of the 2003 survey and show that PROGRESA has significant impacts on the probability of stable employment immediately following high school completion among disadvantaged children (proxied by rainfall shocks in early childhood) but no impact on children with greater endowments. Parker & Vogl (2018) use a difference-in-differences strategy and leverage both the spatiotemporal variation in program roll-out at the municipal level and cohort variation in the age at which children were treated. They find that childhood exposure to PROGRESA improves educational attainment, geographic mobility, labor market outcomes, and household economic outcomes in early adulthood: The program increased mean labor force participation by 30–40% and labor income by 50% for women. Kugler & Rojas (2018) exploit similar sources of variation, combined with propensity score weighting, and estimate significant positive impacts of the program on both the likelihood and quality of employment.

In contrast, Molina Millán et al.'s (2018) evaluation of the Honduras Programa de Asignación Familiar phase II (PRAF II) conditional cash transfer program estimates more ambiguous effects on beneficiaries' labor market outcomes. They use national census microdata and assign individuals their treatment status based on their municipality of birth, the unit of randomization in the RCT. Transfer recipients have significantly higher schooling attainment 13 years after the start of the program, with a notable increase in the likelihood of attaining university studies. Program receipt also more than doubles the probability of international migration among young men. However, impacts on labor market outcomes are less clear cut: Molina Millán et al. find no significant treatment effects on wages or earnings, except for some negative effects on women's hours worked. The labor market results are difficult to interpret, as some young adults are still transitioning into the labor market (for instance, the students enrolled in university), and thus further follow-up surveys could be useful to more reliably assess impacts on lifetime earnings.

There is some evidence that the mode of delivery for cash transfers may be important in determining schooling and other outcomes. For instance, Barrera-Osorio et al. (2017)⁹ leverage administrative data to analyze the Colombian Conditional Subsidies for School Attendance (Subsidios Condicionados a la Asistencia Escolar) program in 2005. The experiment has three treatment arms: the basic bimonthly transfers; the savings treatment, where families are forced to save a portion of the transfers until they make school enrollment decisions; and a tertiary transfer that is conditional on tertiary school enrollment. While the various cash transfer arms are all effective in boosting short-run secondary school enrollment, only the savings treatment improves longer-term educational outcomes, particularly tertiary enrollment.¹⁰

⁷We omit PROGRESA/Oportunidades from **Table 1** because the follow-up studies that are based on the original randomization only report medium-run outcomes.

⁸We omit these studies from **Table 1** because they do not rely on RCT-based estimators.

⁹We omit this study from **Table 1** because it does not report labor market outcomes.

¹⁰We refer interested readers to the contemporaneous work by Molina Millán et al. (2019), which also summarizes and evaluates research on conditional cash transfers, while also bringing in more evidence from non-experimental studies and projects focusing mainly on schooling outcomes.

2.1.4. Scholarship programs. Scholarship or school voucher programs are closely related to some conditional cash transfer interventions in that the award is conditional on school attendance (although they are not identical, since scholarship funding can only be spent directly on education). There are now several long-run RCT evaluations of scholarship programs. As we survey in this section, these tend to show both meaningful gains in educational attainment and subsequent benefits in the labor market.

Bettinger et al. (2018) evaluate the Programa de Ampliación de Cobertura de la Educación Secundaria (PACES) voucher program in Colombia. The program used a lottery to assign vouchers for private secondary schools among applicants from public elementary schools in the poorest two socioeconomic strata in Colombia. The authors sample the lottery winners and losers in 1994 and match their identifications to five different administrative data sets, including a rich set of educational, financial, and labor market outcomes. Up to 20 years later, when applicants' average age was roughly 33, the voucher winners had completed significantly more tertiary education, had experienced lower teen fertility, had annual formal earnings that were 8% higher, and had greater access to formal consumer credit and better credit scores than lottery losers. Notably, these impacts on formal sector earnings are entirely driven by applicants to vocational (as opposed to academic) schools. Effects on formal earnings and payroll taxes are also concentrated among the top 40% of the sample distribution: Much of the voucher effect appears to work through increasing the odds that winners make it into the middle class. A fiscal calculation based on impacts on formal sector earnings and payroll taxes shows that the program is likely to generate large and positive public finance benefits.

Duflo et al. (2018) evaluate a 2008 secondary school scholarship program in Ghana and also find positive effects on some labor market outcomes. The program randomized full scholarships for public high schools among rural youth who had gained admission but did not immediately enroll. Five years after receipt of the scholarship, winners showed increased educational attainment and improved cognitive skills and also engaged in more preventative health behaviors; sample females were less likely to have become pregnant. Nine years after the program, treated individuals were significantly more likely to have public sector jobs, which tend to be characterized by a high wage premium, more benefits, and greater job security. They were also more likely to have jobs with benefits or jobs that they characterized as permanent. Yet there were no significant differences between scholarship winners and losers in total earnings, log earnings conditional on positive earnings, or hours worked for those observed to be working. These null effects on earnings should be interpreted with caution, however, as the confidence intervals are fairly wide and cannot exclude either zero effects or very large private returns. Moreover, a nontrivial portion of participants were still receiving tertiary education, and that portion is significantly higher in the treatment group, raising the possibility that treatment effects may grow over time. This pattern is particularly important if we believe that, as Bettinger et al. (2018) show, gains from these programs tend to be concentrated at the top of the distribution. Duflo et al. caution that, while these results indicate positive private returns to education to the extent that education simply helps people get access to jobs with rents, it may not generate similar social returns. It is also worth noting that the impact of the scholarships on obtaining public sector jobs only became apparent over time, likely because many of the positions require tertiary education.

Taken together, the findings from the existing long-run randomized evaluations of both conditional cash transfer programs and scholarship programs indicate that very long follow-up periods—often of longer than a decade—may be necessary to confidently estimate program impacts on lifetime earnings. This is due to the fact that many beneficiaries are still in school in their 20s, and that certain positions (such as public sector or formal private sector jobs) have rising wage profiles that only become apparent over time. These issues may be particularly important

for females in low-income countries, many of whom also have lower labor market participation in early adulthood than they exhibit later on in life.

A lack of statistical power is also often a challenge in long-term impact evaluation. Income is typically measured with considerable noise, especially in low- and middle-income settings. Measurement concerns are exacerbated by the fact that actual income is highly skewed; that most people in low-income countries obtain a large proportion of their income from self-employment or informal activities, often with strong seasonal variability; and that these income sources may be subject to important reporting biases. For these reasons, in some settings, other socioeconomic indicators, such as jobs with benefits, permanent jobs, and public sector jobs, as in the work of Duflo et al. (2018), may sometimes be more informative about long-run living standards than snapshot income measures. Increasing the frequency of measurements may also be helpful in averaging out measurement errors when dealing with such noisy outcomes (McKenzie 2012).

Finally, we highlight concerns regarding the file drawer problem and publication bias in assessing the studies surveyed in this review. It is possible that studies that deliver null results or less interesting findings are less likely to be published or even to be written up in the first place. Even within published studies, there may be concerns that outcomes with statistically significant results are emphasized over potentially more meaningful outcomes where impacts are less pronounced. This is a particularly important issue given the latitude that researchers often have in selecting results to report across a range of outcomes or sample subgroups. In the sections below, we discuss the importance of collecting comprehensive follow-up measurements across a wide range of experiments, ideally with prespecified outcomes and statistical tests, which could be expected to deliver a more complete picture about the overall impacts of a particular intervention and of the body of evidence as a whole.

2.2. Long-Run Impacts of Child Health Interventions

The literature on the short-term impacts of child health interventions is vast, spanning public health, economics, education, psychology, and nutrition. The RCT evidence on long-run economic impacts of health is far more limited. The limited existing evidence finds generally positive impacts of child health interventions on adult productivity.

2.2.1. Deworming. As discussed in Section 1, the more traditional approach to studying long-term impacts is leveraging historical natural experiments for (hopefully) quasirandom variation in treatment. In the deworming case, Bleakley (2007) studies the successful eradication of hookworm disease from the US South and finds large positive long-run educational and socioeconomic impacts. These results are echoed by experimental evidence in developing countries. The Kenya deworming study (Miguel & Kremer 2004) evaluates an experiment starting in 1998 that randomized 75 schools into an intervention group, with free deworming drug treatment and worm prevention health education, and control groups. The control group schools were phased into deworming treatment two to three years after the early treatment groups, a larger gap between early treatment and late treatment groups than was observed in the experimental PROGRESA/Oportunidades evaluation, for instance. In the short-run, Miguel & Kremer (2004) estimate increased school participation rates and reductions in worm infections among those who directly received drugs, as well as evidence for treatment externalities, but no significant improvements in students' academic or cognitive test scores.

There have since been multiple follow-up survey rounds of the a representative subsample of the deworming sample, in what is called the Kenya Life Panel Survey (KLPS), starting in 2003. These panel (longitudinal) surveys have been characterized by relatively high respondent

effective tracking rates (ETRs)¹¹ of approximately 83.9% (among those still alive), with tracking rates balanced across the treatment and control groups. In the second follow-up round (KLPS-2), collected during 2007–2009, roughly 10 years after the start of the deworming project, Baird et al. (2016a) find that deworming program beneficiaries showed increased educational attainment, especially among women (women were 25% more likely to have attended secondary school), while labor supply increased among men (men worked 17% more hours each week), with accompanying shifts in labor market specialization. Since the deworming treatment is inexpensive (at less than US\$1 per person per year), the authors estimate a large annualized financial internal rate of return of 51.0% when accounting for health spillovers.

There is new evidence of similarly large impacts on economic productivity and living standards in the third KLPS follow-up survey round (KLPS-3), which was collected during 2011–2013, approximately 15 years after the start of the primary school deworming project. Baird et al. (2018) show that respondent tracking rates were similarly high, at 84%, and once again balanced across treatment and control groups. Treatment group respondents still had higher total earnings, with an average gain of 13%, which once again implies an extremely high rate of return to school-based deworming program spending. This KLPS round also featured a detailed consumption expenditure module, which allows for more reliable assessment of household living standards. The data indicate that consumption is also significantly higher in the treatment group, with an average effect of 23%. The gains in both total earnings and consumption are considerably larger among males, echoing results from KLPS-2. Beyond economic productivity and living standards, treatment group beneficiaries are significantly more likely to live in a city than members of the control group, have improvements in certain home characteristics (including improved flooring and greater likelihood of being connected to the electricity grid), and also show gains in subjective wellbeing, specifically in their answers to a question about happiness. Taken together, the Kenya deworming project provides evidence of meaningful long-run gains in economic productivity and living standards along multiple dimensions at both 10 and 15 years following the start of the intervention.

While the evidence on the benefits of deworming on the labor market comes primarily from the Miguel & Kremer (2004) sample, there is evidence on deworming's educational and cognitive impacts in a related sample. Ozier (2018)¹² estimates large cognitive gains 10 years after the start of treatment among children who were 0–2 years old when the Kenya deworming program was launched and who lived in the catchment area of a treatment school. These children were not directly treated themselves but were in position to benefit from positive within-community externalities generated by mass school-based deworming. Ozier (2018) estimates average test score gains of 0.3 standard deviation units, which is equivalent to roughly half a year of schooling. It is worth noting that the Baird et al. (2016a) sample (who were already enrolled in primary school) did not experience improvements in test scores, which is consistent with the hypothesis that nutritional interventions are particularly effective in improving child cognition in critical early periods. These patterns among two distinct samples across multiple time points taken together indicate that the treatment effects found by Ozier (2018), Miguel & Kremer (2004), and Baird et al. (2016a, 2018) are unlikely to be driven by chance.¹³

As we discuss below, many other studies show that early childhood interventions in utero or before age 3 can have large positive impacts (Gertler et al. 2014, Hoddinott et al. 2008). Evidence

¹¹The ETR is a function of the regular phase tracking rate (RTR) and intensive phase tracking rate (ITR) as follows: $ETR = RTR + (1 - RTR) \times ITR$.

¹²We omit this study from **Table 1** because it does not report labor market outcomes.

¹³Croke & Atun (2019) similarly estimate long-run deworming impacts on test scores in Uganda, and this study is also excluded from **Table 1**, since it does not report labor market outcomes.

from the Kenya deworming project suggests that health interventions among somewhat older school-aged children can also have sizable long-run impacts on labor market outcomes through a combination of impacts on education, nutrition, and health status.

2.2.2. Nutritional supplementation. The earliest experimental evidence on long-run returns to child health interventions comes from the well-known Institute of Nutrition of Central America and Panama (INCAP) experiment in rural Guatemala. Between 1969 and 1977, two nutritional supplements—a high-protein energy drink versus a low-energy drink devoid of protein—were randomly assigned to preschool children in four villages (Behrman et al. 2009, Hoddinott et al. 2008, Maluccio et al. 2009). Researchers find evidence of a 46% gain in adult wages for males who were exposed to the nutritional supplement before 3 years of age (Hoddinott et al. 2008). They also find improved cognitive skills among both men and women (Maluccio et al. 2009) and even some positive intergenerational effects on the nutrition of the female beneficiaries' children up to 35 years later. This is a highly unusual and exceptional data collection effort, and it provides evidence that childhood health and nutrition gains can have large returns in terms of adult labor productivity.

Through the lens of more recent studies, the pioneering INCAP study also has some limitations. First, it has a small effective sample size of just four villages (since the intervention did not vary within villages), and not all of the existing studies fully account for the intraclass correlation of respondent outcomes in their analyses, thus perhaps leading them to overstate the statistical significance of estimated effects. Second, within each village, receipt of the nutritious drink was voluntary, so those who were treated were not a random sample of the population within each village. In this case, the most convincing estimation strategy may be an intention to treat analysis, yet some studies report the direct effects of receiving nutritious drinks on outcomes, potentially introducing selection bias. Finally, sample attrition is a concern in both the 1988–1989 follow-up and the most recent surveys, as more than one-quarter of the original sample was apparently lost by 1988–1989, and roughly 40% was lost by the time of the 35-year follow-up survey.

A public health study by Prado et al. (2017)¹⁴ follows up on the sample from a more recent experiment, the Supplementation with Multiple Micronutrients Intervention Trial (SUMMIT), which provides maternal supplementation with multiple micronutrients (MMN) or iron and folic acid (IFA) in Indonesia. The MMN intervention provided the same nutrients as IFA, plus various vitamins, zinc, copper, selenium, and iodine, which are thought to have benefits for development in utero. The project had a massive sample size of 31,290 women enrolled in the trial during 2001–2004. The authors find that the children (who were 9–12 years old at the time of the follow-up survey) had better cognition and academic achievement if their mothers had been assigned to MMN instead of IFA. This opens up the possibility of longer-term labor market gains, although these are yet to be established in this sample. Unfortunately, as with the INCAP study, sample attrition in the SUMMIT sample is substantial: Only 62% of participants were re-enrolled in the follow-up, among which a representative subset of children was selected for cognitive testing.

2.2.3. Cognitive stimulation. The well-known Jamaica experiment (Gertler et al. 2014) carried out during 1986–1987 provides some of the earliest and most compelling evidence on the long-run benefits of early childhood psychosocial stimulation in a low-income country. The intervention targeted growth-stunted toddlers and consisted of weekly visits over a two-year period by community health workers who taught parenting skills and ways to interact with children to develop cognitive and socioemotional skills. The authors find that, 20 years later, the intervention

¹⁴We omit this study from **Table 1** because it does not report labor market outcomes.

increased participants' full-time job earnings by a massive 25%. For nontemporary jobs, the gains are even higher, at 48%.

These labor market gains could result from increased parental investments in children, increased schooling, and migration. At the end of the two-year intervention, the researchers find that the treatment increased the quality of parental interaction and investment in children, as measured by the HOME inventory (Caldwell & Bradley 1984). These effects faded out in mid-to late childhood (at age 7 and later at age 11), but then ultimately translated to more years of schooling attainment, again illustrating that the absence of effects at one time point does not preclude finding effects later. The authors also find suggestive evidence that the treated group tended to migrate more, and that migrants earned substantially more than those who stayed in Jamaica.

The Jamaica study achieved a fairly low attrition rate of 18.6%, which is much lower than several other early experiments described in this section, including the 40% attrition in the INCAP experiment and 49% in the Turkish Early Enrichment Project (TEEP), discussed below. One important limitation, however, is its modest sample size of 129. Another caveat is that the authors were only able to track 14 out of 23 migrants in the sample, and treatment group individuals were overrepresented among the 14 migrants tracked. This differential attrition of migrants across treatment arms could potentially bias treatment effect estimates upward.

Despite the large positive gains to small-scale psychosocial stimulation programs, some efforts to scale up these interventions have been less successful. Andrew et al. (2018) study a scalable psychosocial stimulation intervention implemented at larger scale than and using the institutional infrastructure of existing government services. Two years after the program ended, they find no effects on child test scores, cognition, behavior, stimulation in the home environment, or maternal depressive symptoms. The authors note that it is possible that intervention effects may appear later on, and long-term effects are unknown.

Another early RCT in the psychology literature, the TEEP (Kagitcibasi et al. 2009), provides further evidence on an early childhood stimulation intervention carried out during 1983–1985 among children aged 4–6 from deprived backgrounds. The intervention randomized children into one of three alternative care environments: an educational day care center, a custodial day care center, or the home. Half of the mothers in each care environment were randomly assigned to receive parenting training related to cognitive stimulation. The 22-year follow-up analysis grouped all treatment arms together into any stimulation and found that treated participants had more favorable outcomes in terms of educational attainment; occupational status; and integration into modern urban life, such as owning a computer. The effects of the enrichment treatment on consumption were positive but not statistically significant. Two limitations are the high sample attrition rate of 49% mentioned above and the fact that assignment to the different preschool environments was not entirely random but was determined in part by availability at the workplace, possibly leading to some selection bias.

2.2.4. Perinatal interventions. There are many RCTs involving perinatal interventions in public health, but they have received relatively little attention from economics researchers to date. While it is unusual for public health studies to collect long-run employment and labor market outcomes, Charpak et al. (2016) do so. They study the 20-year impacts of a kangaroo mother care (KMC) intervention¹⁵ in Colombia and find that the intervention increased beneficiaries' school attendance and, later, wages and labor force participation. However, sample attrition was

¹⁵KMC is an intervention designed for preterm and low-birth-weight infants consisting of (a) continuous skin-to-skin contact between mother and infant, (b) exclusive breastfeeding when possible, and (c) timely (early) discharge with close follow-up (Charpak et al. 2016).

again substantial, unfortunately: The authors were only able to survey 441 participants (62% of the original participants), including 264 participants weighing less than 1,800 g at birth, who were thought to be most likely to gain from the intervention. Another potential methodological concern is the fact that statistical significance levels were not adjusted for multiple hypothesis testing.

The Promotion of Breastfeeding Intervention Trial (PROBIT) in Belarus randomized 31 maternity hospitals and affiliated polyclinics to either the control arm or the intervention, which aimed at increasing breastfeeding duration and exclusivity, during 1996–1997. In a survey carried out 16 years later, Martin et al. (2017)¹⁶ successfully follow up 79.5% of the 17,046 breastfeeding mother–infant pairs who participated in the original trial. They do not find any effects of the intervention on the obesity or blood pressure levels of the infant beneficiaries (who were young adults at the follow-up survey). However, it remains an open question whether this intervention impacts other health outcomes or any cognitive and economic outcomes in the long run.

Baranov et al. (2017) evaluate an intervention that provided psychotherapy to perinatally depressed mothers in rural Pakistan. The intervention successfully reduced depression at the time. Seven years later, it also increased women's financial empowerment, control over household spending, and time- and monetary-intensive parental investments, especially on girls. These investments have the potential to translate into later gains in cognition, education, and labor market outcomes, although longer-term effects are unknown.

2.3. Differential Impacts by Gender

Substantial heterogeneity in treatment effects along gender lines is common across several of the interventions that we survey in this article. However, the literature does not seem to converge on whether it is men or women who consistently gain more from the interventions, or on the mechanisms driving the differences. In this section, we highlight the findings from the long-run studies that we review in **Table 1** and call for further research to help reconcile these findings with each other, as well as with predictions from economic theory.

Baird et al. (2016a, 2018) observe that school deworming treatment effects in both total earnings and consumption are larger in magnitude among males 10 to 15 years after the intervention, although differences are not always statistically significant. In contrast, women who were eligible for deworming as girls are 25% more likely to have attended secondary school, halving the gender gap, and they reallocate time away from traditional agriculture and into cash crops and entrepreneurship. Men who were eligible as boys stay enrolled for more years of primary school, work 17% more hours each week, spend more time in entrepreneurship, are more likely to hold manufacturing jobs, and miss one fewer meal per week (Baird et al. 2016a). The authors argue that these results are broadly consistent with the theory of human capital presented by Pitt et al. (2012), in which time allocation depends on how the labor market values both improved human capital and improved raw labor capacity, and this may vary by gender in low-income brawn-based economies. In particular, Pitt et al. (2012) present evidence consistent with a model in which exogenous health gains tend to reinforce men's comparative advantage in occupations requiring raw labor, while leading women to obtain more education and move into more skill-intensive occupations.

Unlike the case of primary school deworming, Barham et al. (2017, 2018) find that a conditional cash transfer program in Nicaragua generated similar effects on earnings and labor market participation for both men and women, and they uncover quite different underlying causal mechanisms that in many ways are the reverse of those identified in the KLPS. Unlike the case of deworming, both education and learning gains in this case are concentrated among males, in a context where

¹⁶We omit this study from **Table 1** because it does not report long-run labor market outcomes.

boys typically drop out of school at younger ages than girls. Women experienced at most modest effects on education and learning, but improved nutrition and reproductive health during teenage years, which the authors argue could translate into labor market gains.

Duflo et al. (2018) study the impacts of secondary school scholarships in rural Ghana and find larger effects on learning and progress to tertiary education among females. In particular, they note that the marginal males (who were only sent to secondary school because of the scholarship) were much less likely to go on to tertiary education than inframarginal males, while marginal females were just as likely to go on to tertiary education as inframarginal females. They argue that families may typically already send academically promising boys to senior secondary schools even in the absence of scholarships, but that there may be heterogeneity among households in their treatment of girls, with some but not all households sending promising girls to school in the absence of scholarships. The marginal girls could therefore have higher underlying ability than similarly marginal boys.

Bettinger et al. (2018) examine the effects of private secondary school scholarships in Colombia and observe large positive effects on the probability of ever enrolling in tertiary education (including vocational schools and universities), formal credit access, and formal sector earnings, with the strongest scholarship impacts among vocational school applicants, as noted above. Within the vocational subpopulation, there are larger effects among males; within the academic subpopulation, females seem to benefit more, although differences across gender tend not to be statistically significant.

Several other studies find more positive long-run effects for males. Hoddinott et al. (2008) find that, in the INCAP experiment (the nutritional intervention in Guatemala), all effects are concentrated among men, and effects for women are typically smaller and not statistically significant. Molina Millán et al. (2018) find that the conditional cash transfer in Honduras led to increased international migration for young men, by 3 to 7 percentage points, with the effects being smaller for women. There is also evidence that, among those who received the conditional cash transfers, women, but not men, reduced their labor supply. The authors caution that this does not necessarily imply negative labor market impacts for women, as the beneficiaries are still transitioning from school into the labor market. Finally, Blattman et al. (2018b) find that the treatment effects of an entrepreneurial grant in Uganda have largely faded out after nine years. However, among the few impacts that persisted, effects on durable asset ownership were higher among men, whereas effects on occupational choice (such as engagement in a skilled trade) were higher among women.

Further theoretical and conceptual work will likely be needed to make sense of these findings by gender, and additional empirical research will be important to understand which patterns are robust across settings. It will be particularly useful to follow effects over a longer time period and to relate any differences to patterns of marriage, fertility, and female labor force participation across study environments, as well as to patterns of occupational segregation and gender wage gaps.

3. WHAT CAN WE LEARN? OPPORTUNITIES AND LIMITATIONS

The large number of experimental cash transfer and child health studies conducted during the late 1990s and the 2000s provides an opportunity to conduct long-term follow-up studies, as described in Section 1. However, how feasible is this opportunity in practice? In this section, we systematically survey and evaluate the opportunities and limitations of the existing pool of cash transfer and child health RCT studies.¹⁷

¹⁷The overall screening strategy was carried out as part of the Long-term Impact Discovery (LID) project financed by GiveWell and cochaired by Prashant Bharawadj and Craig McIntosh (University of California,

3.1. Cash Transfers

We focus in this section on unconditional or conditional cash transfer experimental studies that examine impacts on either the living standards or economic productivity of individuals and households.

3.1.1. Study screening criteria. **Supplemental Appendix A** provides a detailed description of the screening procedure and justifications for our selection criteria. Study selection was based on six main criteria; namely, for inclusion a study had to

1. have randomized treatment,
2. have been implemented before 2010 (to allow for long-run follow-up),
3. have sufficient statistical power (and, relatedly, a sufficiently large sample size),
4. be properly implemented (in ways that we make precise in **Supplemental Appendix A**),
5. have sufficient differential exposure to the intervention across treatment arms, and
6. have the potential for a reasonably high respondent tracking rate.

Among the 170 publications extracted from the seven meta-analysis studies identified during our review, 19 cash transfer studies appear eligible for long-term follow-up research (see **Table 2**). If we additionally exclude the six studies that have already benefited from a long-term follow-up of labor market outcomes, then 13 studies appear to be particularly promising for new long-term studies. We think of these 13 studies as low-hanging fruit for the research community. Yet the fact that the majority of existing cash transfer RCTs end up being excluded due to important design or data limitations also indicates that many past experiments have, unfortunately, not been set up to allow for longer-term evaluation. In Section 4, we discuss several approaches that could improve this yield rate for future experiments.

3.1.2. Eligible studies. **Table 2** describes the 19 RCTs that meet all of the selection criteria and are considered attractive for conducting a long-term follow-up study. Among these experiments, four—namely, Atencion a Crisis (AAC), Subsidios Condicionados a la Asistencia Escolar (SCAE), Nahouri Cash Transfer Pilot Project (NCTPP), and Zomba Cash Transfers (ZOMBA)—present particularly favorable features: All had interventions that were well implemented, none featured a phase-in design, and none have been the subject of a long-term follow-up survey. Three of these RCTs (namely, NCTPP, ZOMBA, and Tayssir) feature both an unconditional cash transfer study arm and a conditional cash transfer arm, presenting a particularly fruitful setting for comparing the long-run impacts of conditional cash transfers and unconditional cash transfers.

Table 2 also provides two important pieces of information about the selected studies that may guide future decisions regarding whether to conduct a long-term follow-up. First, in the column titled “Phase-in design,” we document whether the original control group subsequently received treatment. Although phase-in studies with sufficient time lag between early treatment and late treatment groups should not be excluded a priori, following up on phase-in studies with a relatively short lag presents some challenges for both estimation and interpretation. We discuss this issue in Section 4.1.

Second, we also report on the short-term impacts of each intervention on the living standards, education, health, and labor market outcomes of household adults (see **Table 2**, columns

San Diego). We thank both of them for their leadership in the project and their crucial intellectual contribution to this section of the paper. The LID project does not focus on education interventions, but in our view, there are also likely to be abundant opportunities for conducting long-term impact evaluations in education given the large number of education RCTs. Assessing the existing pool of education RCTs is beyond the scope of this article.

Table 2 Selected cash transfer studies for potential long-term follow-up

	Study acronym	Country	Type	Start of intervention	Phase-in design ^a	Already followed up (> 5 years)	Short-Term Impacts			
							Economic	Education	Health	Adult labor market
1	PROGRESA (Behrman et al. 2005)	Mexico	CCT	1998	Yes	Yes	+	+	+	0
2	PRAF II (Galliani & McEwan 2013)	Honduras	UCT	2000	No	No	NA	+	NA	0
3	RPS (Maluccio & Flores 2005)	Nicaragua	CCT	2000	Yes	Yes	+	+	+	–
4	BDH (Passon & Schady 2010)	Ecuador	UCT	2003	Yes	Yes	NA	+	+	NA
5	PAL (Cunha 2014)	Mexico	UCT	2003	No	No	+	NA	+	NA
6	SCAE (Barrera-Osorio et al. 2011)	Colombia	CCT	2005	No	Yes	+	+	NA	NA
7	AAC (Macours et al. 2012)	Nicaragua	CCT	2005	No	No	+	+	+	NA
8	YOP (Blattman et al. 2013)	Uganda	UCT	2006	No	Yes	NA	NA	NA	NA
9	MDICP (Kohler & Thornton 2011)	Malawi	CCT	2006	No	No	NA	NA	0	NA
10	BRAC TUP (Bandiera et al. 2017)	Bangladesh	UCT	2007	Yes	Yes	+	NA	NA	+
11	NCTPP (Akresh et al. 2016)	Burkina Faso	Both	2008	No	No	+	+	+	NA
12	Tayssir (Benhassine et al. 2015)	Morocco	Both	2008	Yes	No	+	+	NA	NA
13	ZOMBA (Baird et al. 2011)	Malawi	Both	2008	No	No	NA	+	+	NA
14	Women Plus (Green et al. 2015)	Uganda	UCT	2009	Yes	No	+	NA	NA	+
15	RESPECT (de Walque et al. 2014)	Tanzania	CCT	2009	No	No	NA	NA	+	NA
16	CGP Zambia (Natali et al. 2016)	Zambia	UCT	2010	Yes	No	+	NA	NA	+
17	TASAF (Evans et al. 2014)	Tanzania	CCT	2010	Yes	No	+	+	+	NA
18	BONO (Benedetti et al. 2016)	Honduras	CCT	2010	Yes	No	+	+	+	+
19	Pantawid (Chaudhury et al. 2013)	Philippines	CCT	2009	Yes	No	0	+	+	0

+ indicates significant and positive effects; – indicates significant and negative effects; 0 indicates nonsignificant effects. Abbreviations: AAC, Atención a Crisis; BDH, Bono de Desarrollo Humano; BONO, Bono 10,000; BRAC TUP, BRAC Targeting the Ultra-Poor; CCT, conditional cash transfer; MDICP, Malawi Diffusion and Ideational Change Project; NA, not applicable; NCTTP, Nahouri Cash Transfer Pilot Project; PAL, Programa de Apoyo Alimentario; PRAF II, Programa de Asignación Familiar phase II; PROGRESA, Programa de Educación, Salud y Alimentación; RESPECT, Rewarding Sexually Transmitted Infection Prevention and Control in Tanzania; RPS, Red de Protección Social; SCAE, Subsidios Condicionados a la Asistencia Escolar; TASAF, Tanzania Social Action Fund; UCT, unconditional cash transfer; YOP, Youth Opportunities Program; ZOMBA, Zomba Cash Transfers.

^aFor cases where it is unclear whether there is a phase-in design, we write “No” in this column, but more precisely, this means not to our knowledge.

titled “Short-term impacts”). Conducting follow-up surveys just for studies with large and positive short-term impacts may be tempting, and may even be justified at times, yet focusing solely on these studies can have several undesirable consequences. Cherry-picking only the most favorable studies for follow-up surveys will generate a set of estimated long-term impacts that may be representative of studies that yielded short-run impacts, but would be unrepresentative of the set of studies as a whole. For scientific progress, it would be more useful to conduct follow-up studies for multiple RCTs in **Table 2**, perhaps in a coordinated fashion (with common survey instruments, etc.), to create a more complete picture of long-run impacts.

If one were confident that studies that yielded no short-run impact also had no long-run impact, or that, for example, effects fade out monotonically over time, then one might be able to recover estimates or bounds on long-run impacts more broadly. However, as noted above, there is evidence that the effects of certain development interventions can resurface in the long run even after an apparent fade-out of short-run impacts.¹⁸ The mechanisms underlying this phenomenon are still not well understood. One possibility is that the short- and medium-run surveys fail to adequately capture competencies, such as individual socioemotional skills or job referral networks, that may eventually generate positive impacts. Consequently, failing to follow up samples in which short- to medium-run impacts are modest (or nonexistent) may lead us to erroneously conclude that interventions were unsuccessful when in fact they do improve long-run living standards. A bottom line lesson is that a wide and representative range of studies should be evaluated for long-run impacts, and studies should not be ruled out for long-term follow-up because they do not find economically meaningful or statistically significant short-term impacts.

3.2. Child Health Interventions

The child health literature is even more expansive than the body of cash transfer studies, and its boundaries are less clearly defined. We consider studies that aim to improve the overall health of a child from in utero through adolescence.¹⁹ Our criteria include physical health interventions, as well as psychological stimulation and preschool-age child development interventions. The selection criteria do not include education studies beyond preschool unless the intervention specifically included a health component.²⁰

3.2.1. Study screening criteria. We implement a strategy similar to that employed in the cash transfer literature to identify existing child health RCTs that could potentially benefit from a long-term follow-up study. We identify a total of 378 publications and, based on the same criteria used for cash transfers, restrict the selection to 77 eligible studies; details are provided in **Supplemental Appendix B**. As indicated in **Supplemental Appendix B**, these studies are grouped into five main

¹⁸For instance, Gertler et al. (2014) report a lack of medium-run impacts, and Banerjee et al. (2016) report effects that grow over time. Deming (2009) and Chetty et al. (2011a) show that the Head Start program and the Tennessee STAR experiment in the United States improved participant outcomes in adulthood, despite initial fade-out of test score gains.

¹⁹We focus on interventions that address a public health issue and affect a meaningful proportion of children. For instance, stunting is estimated to impact 24.3% of the children under 5 for less developed regions (UNICEF et al. 2018), and the prevalence of malaria is estimated at 9.13% for low Socio-Demographic Index regions in 2017 (see the Institute for Health Metrics and Evaluation’s Global Health Data Exchange, <http://ghdx.healthdata.org/>). Interventions that aim to address specific syndromes or diseases (such as genetic defects that affect very small proportions of newborns) were thus excluded from our review. More details on the inclusion criteria are provided in **Supplemental Appendix B**.

²⁰As noted above, there is a large pool of education RCTs in development economics, but assessing their suitability for long-run follow-up impact evaluation is beyond the scope of this article.

categories (namely, nutrition, perinatal, sanitation, specific diseases, and stimulation). Studies in the nutrition literature, listed in **Table 3**, represent the largest group (32 studies), while the other categories, as shown in **Supplemental Table B.1**, include 45 studies. In the rest of this section, we focus, for reasons of space, on studies in the nutrition literature and leave a detailed discussion of other categories for **Supplemental Appendix B**.

3.2.2. Eligible nutrition studies. Studies of two interventions clearly stand out in **Table 3**: vitamin A and mixed supplementation. Mixed supplementation includes both MMN supplementation and lipid-based nutrient supplements (LNS), as we discuss in more detail below.

Since the mid-1980s, vitamin A interventions have attracted considerable attention among nutritionists. A seminal study by Sommer et al. (1986) among 480 villages in Indonesia suggests that vitamin A supplementation could be a highly effective strategy for reducing mortality (−34%). Since the mid 1980s, multiple RCTs (for a review, see Sommer 2008) have confirmed the positive effects of vitamin A, with an effect size varying between −50% and −34% (though 2 out of 16 studies found no significant impact on health; see **Table 3**). While a more recent and large-scale study has led to some questions regarding these magnitudes (Awasthi et al. 2013), there remains a broad consensus that vitamin A delivery to vitamin A-deficient children or pregnant woman is likely to be an effective strategy for reducing mortality.

Yet how these early health benefits translate into subsequent motor, cognitive ability, or economic productivity impacts in the long run remains almost entirely unknown, as no such long-run studies based on experimental data exist (to our knowledge). This appears to be a promising area for future research. **Table 3** provides some additional information on the vitamin A studies that could feasibly be followed up today. The data presented in the table appear to confirm that vitamin A's short-term impact on health outcomes, and particularly child mortality, is positive overall.

It is possible that an intervention that affects mortality could pose methodological problems for researchers examining long-run outcomes, due to the possibility of selection (or survivorship) bias. Yet we do not believe that this would be a major concern in practice. In the Indonesia data of Sommer et al. (1986), for instance, mortality amounts to only 1% of the total attrition and 0.2% of the differential attrition in a 13-month follow-up. Thus, we do not believe that concerns about differential mortality across treatment arms should deter researchers from following up on populations that took part in vitamin A RCTs.

Another potential methodological challenge posed by the nutritional supplementation RCTs is imperfect compliance in the control group: Due to ethical concerns, in certain trials, project health staff examined control group participants and opted to provide treatment to control group children with severe nutritional problems. This practice makes estimated treatment effects challenging to interpret and seems likely to dampen estimated effects. We flag studies that follow this approach in **Table 3** and **Supplemental Table B.1**.

Mixed supplementation interventions (namely, MMN and LNS) constitute the second-largest group of nutrition studies that we identified, as listed in **Table 3**. Widespread research interest in MMN appears to be more recent than in vitamin A, with most studies dating back only to the mid-1990s. Many of these studies find short-run evidence that MMN supplementation, distributed early on, positively impacts child motor and cognitive development (Eilander et al. 2009). Prado et al. (2017) even report positive medium-term impacts on cognition at age 9–12, but to our knowledge, the impact of MMN on long-run living standards and labor market outcomes has never been estimated with experimental data, creating another promising opportunity for research. The LNS literature is even more recent (starting in the early 2000s). Most studies estimate large positive short-run impacts of such interventions, with gains even larger than those found for MMN interventions (Matias et al. 2017).

Table 3 Selected child nutrition studies for potential long-term follow-up

	Study	Country	Description	Start of intervention	Clustered RCT	Sample size ^a	Age of children	Short-term impacts	
								Health	Cognition
1	Sommer et al. (1986)	Indonesia	VA	1983	Yes	450	12–71 mo	+	NA
2	Rahmathullah et al. (1990) ^b	India	VA	1985	Yes	206	6–60 mo	+	NA
3	Vijayaraghavan et al. (1990)	India	VA	1987	Yes	84	1–5 y	0	NA
4	Herrera et al. (1992) ^b	Sudan	VA	1988	No	28,753	9–72 mo	+	NA
5	Stansfield et al. (1993)	Haiti	VA	1988	No	11,124	6–83 mo	–	NA
6	Dibley et al. (1996) ^b	Indonesia	VA	1989	No	1,407	6–47 mo	+/–	NA
7	Ross et al. (1993) (VAST) ^b	Ghana	VA	1989	Yes	185	6–90 mo	+	NA
8	Barreto et al. (1994)	Brazil	VA	1990	No	1,240	6–48 mo	+	NA
9	West et al. (1991) ^b	Nepal	VA	1991	Yes	261	6–72 mo	+	NA
10	Shankar et al. (1999)	Papua New Guinea	VA	1995	No	480	6–60 mo	+	NA
11	Jinabhai et al. (2001)	South Africa	VA & deworming	1995	No	579	8–10 y	+	0
12	Sempéregui et al. (1999) ^b	Ecuador	VA	1996	No	400	6–36 mo	+	NA
13	Lind et al. (2004)	Indonesia	Iron & zinc	1997	No	680	6–12 mo	+	+
14	Rahman et al. (2001) ^b	Bangladesh	VA & zinc	1997	No	800	12–35 mo	+/–	NA
15	Solon et al. (2003)	Philippines	MMN & deworming	1998	No	831	6–14 y	+	+
16	Sivakumar et al. (2006)	India	MMN	1999	Yes	20	6–16 y	+	+
17	Awasthi et al. (2013) (DEVTA)	India	VA & deworming	1999	Yes	72	6–72 m	+	NA
18	Suppl. Mult. Micronut. Interv. Trial Study Group et al. (2008) (SUMMIT)	Indonesia	MMN	2001	Yes	262	In utero	+	NA
19	Manger et al. (2008)	Thailand	MMN	2002	No	569	5–13 y	+	+
20	Faber et al. (2005)	South Africa	MMN	2002	No	361	6–12 mo	NA	+
21	Sazawal et al. (2010)	India	MMN	2002	No	1,257	1–4 y	+	NA
22	NEMO Study Group (2007) (NEMO)	Indonesia	MMN & fatty acid	2003	No	384	6–10 y	NA	+
23	Long et al. (2006)	Mexico	VA & zinc	<2005 ^c	No	736	6–15 mo	+/–	NA
24	Aboud et al. (2009)	Bangladesh	Responsive feeding	2007	Yes	37	8–20 mo	0	+
25	Suchdev et al. (2012) ^b	Kenya	MMN	2007	Yes	60	6–35 mo	+	NA
26	Aboud & Akhter (2011)	Bangladesh	MMN & responsive feeding	2008	Yes	45	8–20 mo	+	+
27	Veenemans et al. (2011)	Tanzania	Zinc & MMN	2008	No	612	6–60 mo	0	NA
28	Maleta et al. (2015) (LiNS-DOSE) ^b	Malawi	LNS & milk	2009	No	1,932	5–7 mo	0	NA
29	Aboud et al. (2009) (LiNS-DYAD)	Ghana	MMN & LNS	2009	No	1,320	In utero	+	NA
30	Attanasio et al. (2014)	Colombia	Stimulation & MMN	2010	Yes	96	12–24 mo	0	+
31	Hess et al. (2017) (LiNS-BF)	Burkina Faso	MMN	2010	Yes	36	6–27 mo	+	NA
32	Mazumder et al. (2015) (Neovita)	India	VA	2010	No	44,984	Newborn	0	NA

+ indicates significant and positive effects; – indicates significant and negative effects; 0 indicates nonsignificant effects; +/– indicates coexistence of significant positive and negative effects (including side effects). Abbreviations: LNS, lipid-based nutrient supplement; MMN, multiple micronutrient; NA, not applicable; RCT, randomized controlled trial; VA, vitamin A.

^aWe report the number of clusters for clustered RCTs and the number of households or individuals for nonclustered RCTs.

^bIn these RCTs, participants in treatment and control arms are regularly examined during the trial, and those with severe conditions (e.g., severe VA deficiency) are then treated; this practice may change the interpretation of estimated treatment effects.

^cThe authors did not mention when the intervention was conducted, but we infer that it was before 2005, when the paper was submitted.

Although some appear promising, the bulk of MMN and LNS RCTs are still too recent for a long-term follow-up on economic outcomes and are thus excluded from **Table 3**. However, many will become viable candidates in the coming years.

4. HOW CAN WE DO BETTER? RESEARCH DESIGN AND DATA

This section contains a discussion of the following two questions: (a) How can researchers most effectively assess the long-run impacts of an intervention that has already been conducted, and (b) how can researchers design experiments and data collection to improve the feasibility of studying long-run impacts? We discuss these two intertwined issues in the context of both research design and data collection and usage.

4.1. Research Design

The most important building block of a RCT is the experimental design. One type of design that is common in field experiments in economics, especially among those that we review in this article, is the phase-in design.²¹ A phase-in design is one in which treatment groups first receive the interventions, and control groups then receive the same interventions later.

A phase-in design ensures greater similarity across the treatment and comparison groups in the (eventual) distribution of assistance, arguably relaxing some ethical concerns, and may also increase the local political acceptability of a project. It is also a natural design choice when real-world programs are being piloted or gradually rolled out: Randomizing the order of program expansion generates treatment and control groups. These experiments include many of the earliest and most influential studies in development economics, including some of those that have already carried out long-run follow-ups. Examples of phase-in designs include the prominent PROGRESA/Oportunidades experiment (Parker & Todd 2017), the deworming program in Kenya (Miguel & Kremer 2004), and the graduation programs (Bandiera et al. 2017).

One might be tempted to exclude phase-in experiments when trying to learn about long-term impacts due to concerns that long-term effects are not identifiable when there are no pure control groups left. However, we demonstrate that, under certain assumptions detailed below, it is possible to identify long-term treatment effects in the presence of a phase-in design as long as measurements are taken sufficiently frequently. We also show that the variance of treatment effect estimates will grow linearly over time, at a rate that varies inversely to the difference in the duration of treatment between the treatment and control groups (as denoted by T below).

Consider a setup similar to that of Borusyak & Jaravel (2016), in which a panel of units (individuals or clusters) $i = 1, \dots, 2N$ are randomized into two (equally sized) groups $j = 0, 1$, which are the control group (or late treatment group) and the treatment group (or early treatment group), respectively. Suppose that the treatment group receives the treatment at period $T_0 = 0$, and that the intervention is phased into the control group after $T_1 = T$ periods. First consider a situation in which the outcomes Y_{ijt} are observed $(K + 1)$ times, at $t = 0, T, \dots, KT$ (calendar time). Following the event-study notation, denote relative time to treatment $K_{jt} = t - T_j$.²² This denotes

²¹Phase-in RCT designs—also called stepped-wedge designs—appear to be less common in the health research literature.

²²The setup here ensures that, for both the treatment and the control group, $K_{jt} \in \{0, T, \dots, KT\}$ and take the same set of values. When this is not the case, interpolation is necessary, as is the case for Bandiera et al. (2017). They take measurements in years 2, 4, and 7, and their control group is treated in year 4. They interpolate between 2- and 4-year estimates of effects for the treatment group to derive a counterfactual 3-year effect for the control group, to estimate treatment effects after 7 years.

the amount of time that group j has already been exposed to the intervention at time t . We specify the data-generating process to be

$$Y_{ijt} = \alpha_t + \sum_{k=0}^K \tau_k \mathbf{1}\{K_{jt} = kT\} + \epsilon_{ijt} \quad 1.$$

and make the following assumptions:

Assumption 1 (stable dynamic effects). The pattern of dynamic treatment effects (the τ_k terms) is the same in the treatment group and the control group. This holds if the dynamic treatment effects do not interact with (calendar) time, for example.

Assumption 2 (validity of randomization). Absent the intervention, the outcomes of the units in the treatment and control groups follow the same trends.

Assumption 3 (stable unit treatment value assumption). The intervention on the treatment units does not have effects on the outcomes of the control units.

The first assumption is standard for event-study designs (Borusyak & Jaravel 2016), but it is somewhat restrictive, as discussed below. The second and third are standard assumptions for most RCTs.

Note that this flexible setup imposes minimal assumptions on the dynamics of the treatment effects. The treatment effects can be increasing or decreasing over time and can even reverse signs after a certain period.

Under the three assumptions described above, most importantly Assumption 1, the difference between treatment and control groups before the program rolls out to the control group identifies the effects of the program in the first T periods. These estimates can then be used to compute the counterfactual of the control group (if they had been left untreated) to back out long-term impacts after full program roll-out. The long-term effects at $t = 2T$, for example, would be the sum of the difference between treatment and control groups at $t = T$ and at $t = 2T$; in other words, the counterfactual outcome for the control group at $t = 2T$ is simply its actual value minus the estimated effect T periods after the treatment, which is simply the difference between the treatment and control groups at time $t = T$. With the same logic, the long-term effects at $t = 3T$ would be the sum of the difference between treatment and control groups at $t = T$, at $t = 2T$, and at $t = 3T$. One can extend this to $t = KT$, the completed period for which we have measurements of the outcomes, although intuitively, summing up these treatment effect estimates will lead to larger standard errors as t grows.

An important result is that sufficiently frequent measurement is essential. Identification is possible only if the measurements are carried out at least every T periods; otherwise, one simply cannot identify the effects in the initial few periods and cannot compute longer-term effects using the approach described above. However, in the case where the initial measurement is done after phase-in of the control group, if we are willing to make the assumption that the effects of additional exposure are nonnegative, then the difference between treatment and control groups provides a lower bound of the true treatment effect.

When we run the regression of the form

$$Y_{ijt} = \alpha_t + \sum_{k=0}^K \hat{\tau}_k \mathbf{1}\{K_{jt} = kT\} + \epsilon_{ijt}, \quad 2.$$

we recover the $\hat{\tau}_k$ s. The variance of these estimators is (under standard assumptions, in particular, homoskedasticity)

$$\text{Var}(\hat{\tau}_k) = \frac{2}{N}(k+1)\sigma^2, \quad 3.$$

where σ is the residual standard error of the regression.²³ It is clear that the variance of treatment effect estimates grows linearly over time (namely, as observations are farther from the time of control group phase-in), as opposed to staying relatively stable over time, as would be the case for a non-phase-in RCT design.

Despite reduced precision for (absolute) long-term estimates in a phase-in design, this approach actually yields more precise estimates for the differential effects. These estimates may be of particular interest if one is interested in testing certain hypotheses, such as whether effects grow or fade out over time. This is because these estimates are taken directly from comparing the treatment and control groups at a point in time and are not computed by summing up or differencing multiple estimates. For example, suppose we want to know whether the treatment effect after T periods is the same as the effect after $2T$. In a standard non-phase-in RCT design, one would have to test the equality between the treatment effect estimates in $t = 2T$ versus $t = T$. In a phase-in design, however, one can take the treatment effect estimate at $t = 2T$ directly, yielding more precise estimates than the former method.

Bandiera et al. (2017) employ a related approach. They evaluate an intervention that was phased in to control groups after four years and compute a range of estimates for the treatment effects after seven years. Rather than calculating standard errors using an analog of the procedure above, they check for robustness by using the 25th, 50th, and 75th percentile quantile treatment effect estimates on the three-year effects to create counterfactuals for the phased-in controls seven years after the program had started. As they measure two-year and four-year effects in practice, they need to impose some additional assumptions for interpolation to get the three-year effects. Adjusting the standard errors with our calculation above leads to somewhat wider confidence intervals than with their approach; one can reject the hypothesis of no long-term effects, however, so the results remain robust under the approach outlined in this article. Note that the phase-in design allows them to demonstrate that effects are in fact increasing over time, even though standard errors on the seven-year effect are fairly large.

While the economics literature generally assumes that the path of dynamic effects does not vary with time (Borusyak & Jaravel 2016), in many contexts, the dynamic path of treatment effects would vary with either the age of participants or other factors that are time varying, such as the prevalence and intensity of a disease. Identification of long-run effects will still be possible if there is a sufficient sample size and sufficient variation in child age (or prevalence and intensity of a disease) among the treatment and control samples to separately identify the dynamic path of effects for children of different ages (or in contexts with different prevalence and intensity of the disease). However, this would be impossible if time and age (or prevalence and intensity) are perfectly correlated. For example, if all of the treatment and control group individuals are four years old at $t = 0$ when the treatment group receives a health intervention, and if the control group receives the intervention at $t = 3$ (three years later), then to estimate long-term effects, we would have to make the perhaps implausible assumption that effects on 4-year-olds are the same as effects on 7-year-olds.

While we show that long-run effects may be econometrically identified even with phase-in designs, they will at best be estimated with more noise, and so our view is that experimental research

²³The derivation is shown in **Supplemental Appendix C**.

designs with pure control groups are generally preferable to phase-in designs when it is ethically and politically feasible to use them.

The other basic building block of a RCT is an adequate sample size. However, many trials are underpowered to detect modest yet economically meaningful treatment effects, partly because researchers often face a trade-off between the number of treatment arms and statistical power. Croke et al. (2016), for example, show that, out of the 22 studies that estimate the impacts of mass deworming, the median sample size for nonclustered RCTs is only 198 individuals, and the median sample size for clustered RCTs is 80 clusters. For assessing long-term impacts, concerns about power are particularly relevant, because sample attrition may further erode statistical power. One may combine data from individual papers and conduct meta-analyses to gain more statistical power and make progress in this area. Study sample size plays a role in our selection criteria, as described above and in **Supplemental Appendices A and B**.

4.2. Data

In the following sections, we discuss several types of data that researchers can utilize for long-term impact evaluation, including household follow-up surveys, administrative data, and various new data sources.

4.2.1. Follow-up surveys. **Table 1** illustrates that follow-up surveys are the most common source of data used to conduct long-run evaluations of RCTs in international development. The choice to use survey data appears to often be made out of necessity: In most low-income countries, relevant administrative data at the individual level are either nonexistent or difficult to obtain. Even when they exist and are accessible, administrative records may only capture a small share of the outcomes of interest to development economists. For instance, few low-income countries rigorously measure informal economic activity, self-employment earnings, or subsistence agricultural production, and even when they do, data may only exist for a small subset of the population (which may not overlap with the population studied in an RCT). It is not a coincidence that the rise in field experiments and original survey data collection in development economics have gone hand in hand over the past 20 years.

Individual- or household-level surveys have many strengths, but also key limitations. The most important upside of original survey data is the researcher's ability to design their own questions to effectively answer the question at hand. Many recent household surveys in development economics collect highly detailed measures of demographic, educational, health, psychological, and labor market and enterprise outcomes. The richness of original survey data, and the fact that questions can be tailored to particular study goals, allows researchers to probe the mechanisms underlying any intervention impacts and explore heterogeneity in treatment effects across subgroups. It has become a rite of passage for young development economists to spend extended periods of time in the field designing and piloting survey questions, improving the implementation of data collection processes, and sitting in on countless surveys with trained enumerators. In our view, a positive byproduct of these real-world experiences is often a better understanding of the study setting.

Two frequent downsides of original survey data collection are cost and attrition. First, relative to the cost of simply downloading existing administrative records, original survey data collection of thousands of respondents is extremely expensive, with typical project budgets running into the hundreds of thousands of dollars. (Of course, downloading relevant administrative data is usually simply not an option in development economics.) Second, follow-up surveys often suffer from considerable sample attrition. As illustrated in **Table 1**, several prominent existing long-run follow-ups feature high attrition rates, including 40% in the INCAP nutritional supplement study, 49%

in the TEEP cognitive stimulation study, and nearly 40% in PROGRESA (Behrman et al. 2011). Sample attrition appears to be particularly severe in settings where migration—both domestic and international—is common, and among adolescent and young adult populations that are particularly mobile geographically as they seek out educational, labor market, and family opportunities.

Fortunately, several recent long-term tracking efforts, such as the Indonesia Family Life Survey (IFLS) (Strauss et al. 2016; Thomas et al. 2001, 2012), the KLPS, and the Ghana study mentioned above (Duflo et al. 2018), report much lower sample attrition rates. These surveys all devote considerable resources to tracking and recontacting original participants, which is critical for reducing nonrandom attrition and improving data quality. To illustrate, the IFLS-5 tracked 92% of the original households after 21 years. This is despite the high geographic mobility of the baseline respondents: In the fourth wave in 2007, over one-third had moved from the community in which they were interviewed at baseline. For KLPS-3, the effective tracking rate is 84% after 15 years and is not significantly different between the deworming treatment and control groups (Baird et al. 2018). Encouragingly, **Table 1** indicates that several recent studies have even higher survey respondent tracking rates over periods of roughly a decade.

How have these projects improved long-term tracking and achieved such low attrition rates? Below, we document several key lessons from the pioneering IFLS project (Thomas et al. 2001, 2012). Several of the authors of this article also have first-hand experience in respondent tracking from KLPS and the Ghana study, and it is also worth stating several lessons that we have learned along the way (Baird et al. 2008).

A first key lesson is that the detailed contact information of the respondent, as well as of their close relatives and neighbors, should be collected as early as possible in the data collection effort. Starting from the first wave, IFLS began collecting the current residential locations of all households, a sketch map with landmarks and a description of how to find the location, land line and mobile phone numbers, email addresses, people who would likely know their whereabouts in the future and their contact information, whether respondents are planning to move and the likely destinations, and so on (Thomas et al. 2012). When tracking respondents, a field team needs as many leads as possible. By the time several years have passed since an intervention started, it may simply be too late to gather this type of data on respondents who are already on the move. Similarly, it is important to renew contact with respondents relatively frequently—in our experience, at least every few years—to prevent residential location information from becoming stale.

Second, respondent tracking has become considerably easier over the past decade or so in many low-income countries as mobile phone penetration has expanded, becoming nearly universal in many societies. At the start of early KLPS follow-up rounds (approximately 15 years ago), launching a tracking round meant revisiting the original villages and schools of the school deworming project; today, a follow-up round is launched with a barrage of cell phone calls and texts to respondents and their relatives to figure out if they have moved and to set up in-person interviews. In the Ghana study, the research team even provided cell phones to respondents at baseline to facilitate later follow-up contacts (although this step may become unnecessary over time as larger shares of individuals own mobile phones). The cost savings and logistical gains for researchers generated by new communication technologies have been immense.

Third, we have observed that respondent tracking in KLPS actually appears to become somewhat easier as respondents age out of their 20s and into their 30s, as many individuals appear to settle into more stable family, work, and residential arrangements. If a panel survey data collection effort can get through the more difficult adolescent and young adult period unscathed, then there is hope for more consistently high tracking rates in midlife and beyond.

Fourth, in many low-income countries, including Kenya, there is substantial mobility across national borders. The KLPS project has always had a policy of tracking respondents who move



internationally, via phone or Skype surveys, if necessary, to limit attrition. While the costs of international tracking can be substantial, it is critical for successful long-run follow-up surveys in many settings. We note that the KLPS survey was launched in a Kenyan region that features a fairly open border with Uganda (and strong family, ethnic, and historic ties across the border), which greatly facilitates both international mobility and international tracking; the situation along other borders may be more challenging, for instance, currently when it comes to the case of Mexican and Central American migrants who have moved to the United States.

Finally, the IFLS research team documents many differences between movers and stayers, including in exhibiting significantly different observed returns to education in IFLS-4 (in 2007) (Thomas et al. 2012). This indicates that treatment effect estimates generated in samples that exclude movers could be biased. Similarly, in the KLPS-3 analysis described above, deworming treatment has substantial positive long-run impacts on the likelihood of urban migration, which suggests that excluding the subsample of movers from the analysis could again lead to bias. Taken together, investing in tracking study respondents across space will likely be valuable for most long-run research projects.

4.2.2. Administrative data. Administrative data can be a highly cost-effective alternative to follow-up surveys in cases where relevant administrative data are available and the baseline surveys contain information that allow them to match to official records (for instance, a government identification number). Bettinger et al. (2018), for example, achieve very high tracking rates among PACES school voucher lottery participants in Colombia, with 97% of participant identification numbers being valid. These individuals can then be matched to five distinct government administrative data sets with minimal attrition and no need for costly follow-up surveys, since the government is already collecting these data. For labor market outcomes, the authors are able to roughly match participants to formal sector earnings and tax payment records in the 2008–2014 SISPRO data set (from Colombia's Social Protection Ministry), as well as to Familias en Accion conditional cash transfer and other social protection program eligibility information in the SISBEN survey. For those living in low-income neighborhoods, they also obtain self-reported earnings. The administrative data approach of Bettinger et al. (2018) is extremely cost-effective and yields a rich set of outcome data.

It seems clear to us that administrative data should be used when high-quality information on relevant measures is available and can be matched to study participants; the key constraint is that this has rarely been the case in practice and is especially rare in the poorest developing countries (note that Colombia is a middle-income country). In assessing the feasibility of additional long-run follow-up projects, researchers could consider the presence of good administrative data, such as in Colombia, as an important criterion.

When there are no unique identifiers in place to help researchers match records in different data sets, using probabilistic matching techniques—matching on individual characteristics such as names, neighborhoods, addresses, birth places, birth dates, etc.—can be an attractive alternative. In Venezuela, Hsieh et al. (2011) match the list of petition signers who opposed the Hugo Chávez regime to household survey respondents to estimate the economic effects of being identified as a Chávez political opponent. Even without an official identification number, the authors successfully match most records based on locality, exact birth date, and gender.

Yet administrative data also have some drawbacks. As mentioned above, administrative records will typically not contain all of the outcomes or measures that researchers are interested in. Where subsistence agriculture and informal sector economic activities are widespread, as in many low-income countries, administrative data will likely miss important components of total household earnings. To some extent, this concern can be ameliorated if there exist proxy means-tested

programs (for poverty alleviation) with accompanying administrative records, but the surveys that are used to determine eligibility may only be collected infrequently or cover limited geographic areas.

Similar strategies are pursued in another Latin American environment by Molina Millán et al. (2018). They evaluate PRAF-II, a conditional cash transfer program in Honduras, using microdata from the national population census and repeated cross-sectional surveys collected more than a decade after the program's start. They assign individual program treatment status based on individuals' municipality of birth in what is essentially an intention to treat design, given that municipalities were the unit of randomization. However, administrative data in this case again have some limitations: The use of aggregated municipal-level data can lead to risk of bias if there is extensive migration and asymmetric mobility across treatment and control areas, for instance.

4.2.3. New data sources. An emerging body of studies has leveraged new data and methods from economics, computer science (specifically machine learning), and earth sciences to measure poverty, and these have some promise. In principle, these methods could offer cost-effective and scalable ways to evaluate international development interventions in a timely manner, especially in cases where original data collection is challenging, such as societies experiencing armed conflict. The key caveat to most of these methods is that they are limited in terms of the outcomes that researchers can examine, falling far short of the richness found in most original development economics household surveys in their measurement of living standards, consumption, and income, and they typically have nothing to say about economically important attitudes, beliefs, and expectations, let alone direct health or nutritional measures.

An early application of new data to estimate RCT impacts is the Alix-Garcia et al. (2013) study, which uses Landsat satellite data to study the ecological consequences of the Mexican PROGRESA/Oportunidades program. Remote sensing data appear particularly well-suited to studying impacts of cluster-randomized interventions (like this Mexico RCT), where treatment and control geographic areas can be easily identified. Researchers may not always have adequate resolution, or relevant geolocation data, to identify treatment and control households when randomization is done within a village. However, there are exceptions: Burke & Lobell (2017) combined high-resolution satellite imagery (1 m Terra Bella imagery) and intensive field sampling on thousands of smallholder maize fields over two years, and they detected positive crop yield responses to fertilizer and hybrid seed inputs (see also Jean et al. 2016). Satellite data have also been used to generate night-light intensity measures, which have recently become very widely used to proxy for overall local economic activity (Henderson et al. 2012); again, these could be useful for the evaluation of RCTs where the unit of randomization is fairly large.

More recently, researchers have begun leveraging cell phone records to assess poverty. The seminal paper by Blumenstock et al. (2015) shows that machine learning methods can be used to predict household wealth and living standards measures from detailed mobile phone metadata in Rwanda. Blumenstock et al. (2018) apply this method to impact evaluation: They recruited mobile phone subscribers in Afghanistan to participate in a 7-month high-frequency phone-based survey and matched their responses to historical call detail records. They are able to infer the onset and magnitude of positive and negative economic shocks, including the (randomized) receipt of cash transfers. In cases where cell phone metadata are available to researchers, baseline survey data collection could usefully collect participants' mobile phone numbers, which could later be matched to call detail records, and together with the appropriate prediction methods, these can generate estimates of living standards.

Yet Blumenstock (2018) also warns that these new data sources may suffer from a lack of validation and biased algorithms. For instance, there is some evidence that existing predictive models

may work in one institutional context but not be nearly as successful in others. The number of international phone calls made, for example, is a better predictor of wealth in Rwanda than it is in Afghanistan. Predictive model performance also appears to deteriorate rather quickly over time, raising questions about how often the models need to be revalidated and at what cost in terms of fresh training data. In addition, the behavioral patterns currently used for prediction may change when individuals become aware that their personal data are being observed and used to generate statistics that affect eligibility for particular government programs, for instance. Moreover, when these predictive models are trained on biased or patchy data, those who are poorly represented (e.g., households too poor to own a smart phone) may be further marginalized, and predictions for important subpopulations may be largely uninformative.

The bottom line on new data sources is similar to that on administrative data: They are cheaper to collect than traditional household surveys and should be used when available, but may lack the specific measures needed to test many important economic research hypotheses. As a result, we do not see original household data collection disappearing from the development economics toolkit anytime soon, including in the context of long-run studies.

5. CONCLUSION

In this article, we argue that the coming years provide an exceptional opportunity for development economists to make intellectual progress in understanding the underlying determinants of long-run living standards by exploiting the large number of development RCTs that have been conducted since the late 1990s. Despite the methodological and data limitations of many early RCTs in development economics and public health, we identify dozens of studies that currently appear amenable to follow-up evaluations, with scores if not hundreds more aging into the possibility of long-run evaluation in the coming decade. If the development economics research community is able to seize this opportunity, then it has the potential to generate considerable scientific progress in our field.

Given the policy relevance and intellectual importance of long-term impact evaluations, we argue that this research agenda should be a top priority for donors and policy makers. Conducting long-term follow-up studies on past RCTs will demand a large amount of funding and coordinated researcher effort to set up successful survey data collection, often across geographic areas and sometimes across academic disciplines. Yet establishing parallel data collection and tracking protocols across multiple interventions could help generate comparable estimates on the long-run impacts of related interventions, leading to greater external validity. There are already models of successful efforts along these lines. Banerjee et al. (2015b), for example, evaluate a multifaceted program targeted at the very poor in six different countries, and a similar effort is underway in the political economy of development through the EGAP Metaketa initiative (<https://egap.org/metaketa>). Comparable long-term evaluations of multiple international development interventions will advance intellectual understanding of the drivers of long-run living standards and could generate valuable insights into comparative cost-effectiveness for policy makers.

We also describe patterns in the relatively small but growing body of literature that already takes advantage of experimental variation to study long-run living standards impacts. One emerging pattern is that several human capital interventions—in both health and education—appear to have successfully led to persistent economic productivity gains, often with impressive rates of return (Baird et al. 2016a). In contrast, most interventions aimed at relaxing liquidity constraints and stimulating firm growth appear to be characterized by positive short-term effects that fade out over time (with the exception of graduation programs that are characterized by large asset

transfers and intensive training and support). This pattern echoes the lack of persistent or meaningful impacts documented in the microcredit literature (see, for example, Banerjee et al. 2015a). Yet we caution that this pattern is driven by a relatively small number of RCT studies and must be viewed as suggestive at this time. With the appropriate resources and coordination, the body of evidence on long-run impacts of these and other development interventions is poised to become much more definitive in the coming years.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We are particularly grateful to Craig McIntosh and Prashant Bharadwaj for their work on the systematic review of cash transfer and child health projects discussed in this article, under the Long-term Impact Discovery (LID) initiative. GiveWell provided generous financial support for the LID initiative, and Josh Rosenberg of GiveWell gave us many helpful suggestions. We are also grateful to David Roodman of Open Philanthropy, who sparked GiveWell's support for the LID project with his 2017 blog post, "How thin the reed?" We also benefited from suggestions and comments provided by the LID faculty advisory committee at the Center for Effective Global Action, including Lia Fernald, Paul Gertler, Marco Gonzalez-Navarro, and Manisha Shah, as well as from Oriana Bandiera, Robin Burgess, Xavier Jaravel, and Rachael Meager.

LITERATURE CITED

- Aboud FE, Akhter S. 2011. A cluster-randomized evaluation of a responsive stimulation and feeding intervention in Bangladesh. *Pediatrics* 127:e1191–97
- Aboud FE, Shafique S, Akhter S. 2009. A responsive feeding intervention increases children's self-feeding and maternal responsiveness but not weight gain. *J. Nutr.* 139:1738–43
- Acemoglu D, Johnson S. 2007. Disease and development: the effect of life expectancy on economic growth. *J. Political Econ.* 115:925–85
- Adhvaryu A, Nyshadham A, Molina T, Tamayo J. 2018. *Helping children catch up: early life shocks and the PROGRESA experiment*. NBER Work. Pap. 24828
- Aizer A, Eli S, Ferrie J, Lleras-Muney A. 2016. The long-run impact of cash transfers to poor families. *Am. Econ. Rev.* 106:935–71
- Akresh R, de Walque D, Kazianga H. 2016. *Evidence from a randomized evaluation of the household welfare impacts of conditional and unconditional cash transfers given to mothers or fathers*. Work. Pap., World Bank, Washington, DC
- Alix-Garcia J, McIntosh C, Sims KR, Welch JR. 2013. The ecological footprint of poverty alleviation: evidence from Mexico's Oportunidades program. *Rev. Econ. Stat.* 95:417–35
- Almond D. 2006. Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 US population. *J. Political Econ.* 114:672–712
- Almond D, Currie J, Duque V. 2017. *Childhood circumstances and adult outcomes: Act II*. NBER Work. Pap. 23017
- Anderson ML. 2008. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *J. Am. Stat. Assoc.* 103:1481–95
- Andrew A, Attanasio O, Fitzsimons E, Grantham-McGregor S, Meghir C, Rubio-Codina M. 2018. Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: a follow-up of a cluster randomised controlled trial in Colombia. *PLOS Med.* 15:e1002556

- Araujo MC, Bosch M, Schady N. 2017. Can cash transfers help households escape an inter-generational poverty trap? In *The Economics of Poverty Traps*, ed. CB Barrett, MR Carter, J-P Chavas, pp. 357–82. Chicago: Univ. Chicago Press
- Attanasio OP, Fernández C, Fitzsimons EO, Grantham-McGregor SM, Meghir C, Rubio-Codina M. 2014. Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in colombia: cluster randomized controlled trial. *BMJ* 349:g5785
- Awasthi S, Peto R, Read S, Clark S, Pande V, et al. 2013. Vitamin A supplementation every 6 months with retinol in 1 million pre-school children in north India: Devta, a cluster-randomised trial. *Lancet* 381:1469–77
- Baird S, Hamory J, Miguel E. 2008. *Tracking, attrition and data quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)*. Tech. Rep., Cent. Int. Dev. Econ. Res., Univ. Calif. Berkeley
- Baird S, Hicks JH, Kremer M, Miguel E. 2016a. Worms at work: long-run impacts of a child health investment. *Q. J. Econ.* 131:1637–80
- Baird S, Hicks JH, Kremer M, Miguel E. 2018. *Worms and wellbeing: 15 year economic impacts from Kenya*. Unpublished manuscript, Univ. Calif., Berkeley
- Baird S, McIntosh C, Özler B. 2011. Cash or condition? Evidence from a cash transfer experiment. *Q. J. Econ.* 126:1709–53
- Baird S, McIntosh C, Özler B. 2016b. *When the money runs out: evaluating the longer-term impacts of a two year cash transfer program*. Unpublished manuscript, Univ. Calif., San Diego
- Bandiera O, Balboni C, Burgess R, Ghatak M, Heil A. 2018. *Why do people stay poor?* Unpublished manuscript, London School Econ.
- Bandiera O, Burgess R, Das N, Gulesci S, Rasul I, Sulaiman M. 2017. Labor markets and poverty in village economies. *Q. J. Econ.* 132:811–70
- Banerjee A, Duflo E, Chattopadhyay R, Shapiro J. 2016. *The long term impacts of a “graduation” program: evidence from West Bengal*. Unpublished manuscript, Mass. Inst. Technol., Cambridge
- Banerjee A, Duflo E, Glennerster R, Kinnan C. 2015a. The miracle of microfinance? Evidence from a randomized evaluation. *Am. Econ. J. Appl. Econ.* 7:22–53
- Banerjee A, Duflo E, Goldberg N, Karlan D, Osei R, et al. 2015b. A multifaceted program causes lasting progress for the very poor: evidence from six countries. *Science* 348:1260799
- Banerjee A, Karlan D, Zinman J. 2015c. Six randomized evaluations of microcredit: introduction and further steps. *Am. Econ. J. Appl. Econ.* 7:1–21
- Baranov V, Bhalotra S, Biroli P, Maselko J. 2017. *Maternal depression, women’s empowerment, and parental investment: evidence from a large randomized control trial*. Discuss. Pap. 11187, Inst. Labor Econ., Bonn, Ger.
- Barham T, Macours K, Maluccio JA. 2017. *Are conditional cash transfers fulfilling their promise? Schooling, learning, and earnings after 10 years*. Discuss. Pap. DP11937, Cent. Econ. Policy Res., Washington, DC
- Barham T, Macours K, Maluccio JA. 2018. *Experimental evidence of exposure to a conditional cash transfer during early teenage years: young women’s fertility and labor market outcomes*. Tech. Rep., Cent. Econ. Policy Res., Washington, DC
- Barrera-Orsorio F, Bertrand M, Linden LL, Perez-Calle F. 2011. Improving the design of conditional transfer programs: evidence from a randomized education experiment in Colombia. *Am. Econ. J. Appl. Econ.* 3:167–95
- Barrera-Orsorio F, Linden LL, Saavedra J. 2017. *Medium- and long-term educational consequences of alternative conditional cash transfer designs: experimental evidence from Colombia*. NBER Work. Pap. 23275
- Barreto ML, Farenzena G, Fiaccone R, Santos L, Assis AMdO, et al. 1994. Effect of vitamin A supplementation on diarrhoea and acute lower-respiratory-tract infections in young children in Brazil. *Lancet* 344:228–31
- Bastian J, Michelmores K. 2018. The long-term impact of the Earned Income Tax Credit on children’s education and employment outcomes. *J. Labor Econ.* 36:1127–63
- Behrman JR, Calderon MC, Preston SH, Hoddinott J, Martorell R, Stein AD. 2009. Nutritional supplementation in girls influences the growth of their children: prospective study in Guatemala. *Am. J. Clin. Nutr.* 90:1372–79

- Behrman JR, Parker SW, Todd PE. 2011. Do conditional cash transfers for schooling generate lasting benefits? A five-year followup of PROGRESA/Oportunidades. *J. Hum. Resour.* 46:93–122
- Behrman JR, Sengupta P, Todd P. 2005. Progressing through PROGRESA: an impact assessment of a school subsidy experiment in rural Mexico. *Econ. Dev. Cult. Change* 54:237–75
- Benedetti F, Ibararán P, McEwan PJ. 2016. Do education and health conditions matter in a large cash transfer? Evidence from a Honduran experiment. *Econ. Dev. Cult. Change* 64:759–93
- Benhassine N, Devoto F, Duflo E, Dupas P, Pouliquen V. 2015. Turning a shove into a nudge? A “labeled cash transfer” for education. *Am. Econ. J. Econ. Policy* 7:86–125
- Bettinger E, Kremer M, Kugler M, Medina C, Posso C, Saavedra J. 2018. *School vouchers, labor markets and vocational education*. Unpublished manuscript, Harvard Univ., Cambridge, MA
- Blattman C, Dercon S, Franklin S. 2018a. *The long run effects of industrial and entrepreneurial jobs: 5-year evidence from Ethiopia*. Work. Pap., Univ. Chicago
- Blattman C, Fiala N, Martinez S. 2013. Generating skilled self-employment in developing countries: experimental evidence from Uganda. *Q. J. Econ.* 129:697–752
- Blattman C, Fiala N, Martinez S. 2018b. *The long term impacts of grants on poverty: 9-year evidence from Uganda's youth opportunities program*. Unpublished manuscript, Univ. Chicago
- Bleakley H. 2007. Disease and development: evidence from hookworm eradication in the American South. *Q. J. Econ.* 122:73–117
- Bleakley H, Ferrie JP. 2013. *Up from poverty? The 1832 Cherokee Land Lottery and the long-run distribution of wealth*. NBER Work. Pap. 19175
- Blumenstock J. 2018. Don't forget people in the use of big data for development. *Nature Comment*, Sept. 10. <https://www.nature.com/articles/d41586-018-06215-5>
- Blumenstock J, Cadamuro G, On R. 2015. Predicting poverty and wealth from mobile phone metadata. *Science* 350:1073–76
- Blumenstock J, Callen M, Ghani T, Keleher N, Shapiro J. 2018. *Measuring poverty and vulnerability in real-time*. Unpublished manuscript, Univ. Calif., Berkeley
- Borusyak K, Jaravel X. 2016. *Revisiting event study designs*. Work. Pap., Harvard Univ., Cambridge, MA
- Brunello G, Comi S. 2004. Education and earnings growth: evidence from 11 European countries. *Econ. Educ. Rev.* 23:75–83
- Buchmann N, Field E, Glennerster R, Nazneen S, Pimkina S, Iman S. 2018. *Power vs money: alternative approaches to reducing child marriage in Bangladesh, a randomized control trial*. Unpublished manuscript, Duke Univ., Durham, NC
- Burke M, Lobell DB. 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *PNAS* 114:2189–94
- Caldwell BM, Bradley RH. 1984. *Home observation for measurement of the environment*. Work. Pap., Univ. Ark., Little Rock
- Chapak N, Tessier R, Ruiz JG, Hernandez JT, Uriza F, et al. 2016. Twenty-year follow-up of kangaroo mother care versus traditional care. *Pediatrics* 139:e20162063
- Chaudhury N, Friedman J, Onishi J. 2013. *Philippines conditional cash transfer program impact evaluation 2012*. Rep. 75533-PH, World Bank, Washington, DC
- Chetty R, Friedman JN, Hilger N, Saez E, Schanzenbach DW, Yagan D. 2011a. How does your kindergarten classroom affect your earnings? Evidence from Project Star. *Q. J. Econ.* 126:1593–660
- Chetty R, Friedman JN, Rockoff JE. 2011b. *New evidence on the long-term impacts of tax credits*. Stat. Income White Pap., Intern. Revenue Serv., Washington, DC
- Croke K, Atun R. 2019. The long run impact of early childhood deworming on numeracy and literacy: evidence from Uganda. *PLOS Negl. Trop. Dis.* 13(1):e0007085
- Croke K, Hicks JH, Hsu E, Kremer M, Miguel E. 2016. *Does mass deworming affect child nutrition? Meta-analysis, cost-effectiveness, and statistical power*. NBER Work. Pap. 22382
- Cunha JM. 2014. Testing paternalism: cash versus in-kind transfers. *Am. Econ. J. Appl. Econ.* 6:195–230
- de Walque D, Dow WH, Nathan R. 2014. *Rewarding safer sex: conditional cash transfers for HIV/STI prevention*. Work. Pap., World Bank, Washington, DC

- Deming D. 2009. Early childhood intervention and life-cycle skill development: evidence from Head Start. *Am. Econ. J. Appl. Econ.* 1:111–34
- Dibley MJ, Sadjimin T, Kjolhede CL, Moulton LH. 1996. Vitamin A supplementation fails to reduce incidence of acute respiratory illness and diarrhea in preschool-age Indonesian children. *J. Nutr.* 126:434–42
- Duflo E, Dupas P, Kremer M. 2018. *The impact of free secondary education: experimental evidence from Ghana*. Unpublished manuscript, Stanford Univ., Stanford, CA
- East CN. 2019. The effect of food stamps on children's health: evidence from immigrants changing eligibility. *J. Hum. Resourc.* In press
- Eilander A, Gera T, Sachdev HS, Transler C, van der Knaap HC, et al. 2009. Multiple micronutrient supplementation for improving cognitive performance in children: systematic review of randomized controlled trials. *Am. J. Clin. Nutr.* 91:115–30
- Evans D, Hausladen S, Kosec K, Reese N. 2014. *Community-based conditional cash transfers in Tanzania: results from a randomized trial*. Work. Pap., World Bank, Washington, DC
- Faber M, Kvalsvig JD, Lombard CJ, Benadé AS. 2005. Effect of a fortified maize-meal porridge on anemia, micronutrient status, and motor development of infants. *Am. J. Clin. Nutr.* 82:1032–39
- Galiani S, McEwan PJ. 2013. The heterogeneous impact of conditional cash transfers. *J. Public Econ.* 103:85–96
- Gennaioli N, La Porta R, Lopez-de Silanes F, Shleifer A. 2012. Human capital and regional development. *Q. J. Econ.* 128:105–64
- Gertler P, Heckman J, Pinto R, Zanolini A, Vermeersch C, et al. 2014. Labor market returns to an early childhood stimulation intervention in Jamaica. *Science* 344:998–1001
- Green EP, Blattman C, Jamison J, Annan J. 2015. Women's entrepreneurship and intimate partner violence: a cluster randomized trial of microenterprise assistance and partner participation in post-conflict Uganda (SSM-D-14-01580R1). *Soc. Sci. Med.* 133:177–88
- Haushofer J, Shapiro J. 2016. The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. *Q. J. Econ.* 131:1973–2042
- Haushofer J, Shapiro J. 2018. *The long-term impact of unconditional cash transfers: experimental evidence from Kenya*. Work. Pap., Busara Cent. Behav. Econ., Nairobi, Kenya
- Heckman JJ. 2006. Skill formation and the economics of investing in disadvantaged children. *Science* 312:1900–2
- Henderson JV, Storeygard A, Weil DN. 2012. Measuring economic growth from outer space. *Am. Econ. Rev.* 102:994–1028
- Herrera MG, Nestel P, Weld L, El Amin A, Mohamed KA, Fawzi W. 1992. Vitamin A supplementation and child survival. *Lancet* 340:267–71
- Hess SY, Peerson JM, Becquey E, Abbeduto S, Ouédraogo CT, et al. 2017. Differing growth responses to nutritional supplements in neighboring health districts of Burkina Faso are likely due to benefits of small-quantity lipid-based nutrient supplements (LNS). *PLOS ONE* 12:e0181770
- Himanshu, Lanjouw P, Stern N, et al. 2018. *How Lives Change: Palampur, India, and Development Economics*. Oxford, UK: Oxford Univ. Press
- Hoddinott J, Maluccio JA, Behrman JR, Flores R, Martorell R. 2008. Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults. *Lancet* 371:411–16
- Hoynes H, Schanzenbach DW, Almond D. 2016. Long-run impacts of childhood access to the safety net. *Am. Econ. Rev.* 106:903–34
- Hoynes HW, Schanzenbach DW. 2018. *Safety net investments in children*. NBER Work. Pap. 24594
- Hsieh CT, Miguel E, Ortega D, Rodriguez F. 2011. The price of political opposition: evidence from Venezuela's *Maisanta*. *Am. Econ. J. Appl. Econ.* 3:196–214
- Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353:790–94
- Jinabhai CC, Taylor M, Coutousdis A, Coovadia HM, Tomkins AM, Sullivan KR. 2001. A randomized controlled trial of the effect of antihelminthic treatment and micronutrient fortification on health status and school performance of rural primary school children. *Ann. Trop. Paediatr.* 21:319–33

- Kagitcibasi C, Sunar D, Bekman S, Baydar N, Cemalcilar Z. 2009. Continuing effects of early enrichment in adult life: the Turkish Early Enrichment Project 22 years later. *J. Appl. Dev. Psychol.* 30:764–79
- Kohler HP, Thornton RL. 2011. Conditional cash transfers and HIV/AIDS prevention: unconditionally promising? *World Bank Econ. Rev.* 26:165–90
- Kremer M. 2003. Randomized evaluations of educational programs in developing countries: some lessons. *Am. Econ. Rev.* 93:102–6
- Kugler AD, Rojas I. 2018. *Do CCTs improve employment and earnings in the very long-term? Evidence from Mexico.* NBER Work. Pap. 24248
- Lind T, Lönnnerdal B, Stenlund H, Gamayanti IL, Ismail D, et al. 2004. A community-based randomized controlled trial of iron and zinc supplementation in Indonesian infants: effects on growth and development. *Am. J. Clin. Nutr.* 80:729–36
- Long KZ, Montoya Y, Hertzmark E, Santos JI, Rosado JL. 2006. A double-blind, randomized, clinical trial of the effect of vitamin A and zinc supplementation on diarrheal disease and respiratory tract infections in children in Mexico City, Mexico. *Am. J. Clin. Nutr.* 83:693–700
- Ludwig J, Miller DL. 2007. Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Q. J. Econ.* 122:159–208
- Macours K, Schady N, Vakis R. 2012. Cash transfers, behavioral changes, and cognitive development in early childhood: evidence from a randomized experiment. *Am. Econ. J. Appl. Econ.* 4:247–73
- Maleta KM, Phuka J, Alho L, Cheung YB, Dewey KG, et al. 2015. Provision of 10–40 g/d lipid-based nutrient supplements from 6 to 18 months of age does not prevent linear growth faltering in Malawi. *J. Nutr.* 145:1909–15
- Maluccio J, Flores R. 2005. *Impact evaluation of a conditional cash transfer program: the Nicaraguan Red de Protección Social.* Discuss. Pap., Int. Food Policy Res. Inst., Washington, DC
- Maluccio JA, Hoddinott J, Behrman JR, Martorell R, Quisumbing AR, Stein AD. 2009. The impact of improving nutrition during early childhood on education among Guatemalan adults. *Econ. J.* 119:734–63
- Manger MS, McKenzie JE, Winichagoon P, Gray A, Chavasit V, et al. 2008. A micronutrient-fortified seasoning powder reduces morbidity and improves short-term cognitive function, but has no effect on anthropometric measures in primary school children in northeast Thailand: a randomized controlled trial. *Am. J. Clin. Nutr.* 87:1715–22
- Martin RM, Kramer MS, Patel R, Rifas-Shiman SL, Thompson J, et al. 2017. Effects of promoting long-term, exclusive breastfeeding on adolescent adiposity, blood pressure, and growth trajectories: a secondary analysis of a randomized clinical trial. *JAMA Pediatr.* 171:e170698
- Martorell R, Habicht JP, Rivera JA. 1995. History and design of the INCAP longitudinal study (1969–77) and its follow-up (1988–89). *J. Nutr.* 125:1027S–41S
- Matias SL, Vargas-Vásquez A, Pérez RB, Valdivia LA, Vivanco OA, et al. 2017. Effects of lipid-based nutrient supplements v. micronutrient powders on nutritional and developmental outcomes among Peruvian infants. *Public Health Nutr.* 20:2998–3007
- Mazumder S, Taneja S, Bhatia K, Yoshida S, Kaur J, et al. 2015. Efficacy of early neonatal supplementation with vitamin A to reduce mortality in infancy in Haryana, India (Neovita): a randomised, double-blind, placebo-controlled trial. *Lancet* 385:1333–42
- McKenzie D. 2012. Beyond baseline and follow-up: the case for more T in experiments. *J. Dev. Econ.* 99:210–21
- Miguel E, Kremer M. 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72:159–217
- Molina Millán T, Barham TCJ, Macours K, Maluccio JA, Stampini M. 2019. Long-term impacts of conditional cash transfers: review of the evidence. *World Bank Res. Obs.* In press
- Molina Millán T, Macours K, Maluccio J, Tejerina L. 2018. *Experimental long-term impacts of early childhood and school age exposure to a conditional cash transfer.* Unpublished manuscript, Paris School Econ.
- Natali L, Handa S, Peterman A, Seidenfeld D, Tembo G, et al. 2016. *Making money work: unconditional cash transfers allow women to save and re-invest in rural Zambia.* Innocenti Work. Pap. 2016-02, Off. Res., UNICEF, New York

- NEMO Study Group. 2007. Effect of a 12-mo micronutrient intervention on learning and memory in well-nourished and marginally nourished school-aged children: 2 parallel, randomized, placebo-controlled studies in Australia and Indonesia. *Am. J. Clin. Nutr.* 86:1082–93
- Ozier O. 2018. Exploiting externalities to estimate the long-term effects of early childhood deworming. *Am. Econ. J. Appl. Econ.* 10:235–62
- Parker SW, Todd PE. 2017. Conditional cash transfers: the case of PROGRESA/Oportunidades. *J. Econ. Lit.* 55:866–915
- Parker SW, Vogl T. 2018. *Do conditional cash transfers improve economic outcomes in the next generation? Evidence from Mexico*. NBER Work. Pap. 24303
- Paxson C, Schady N. 2010. Does money matter? The effects of cash transfers on child development in rural Ecuador. *Econ. Dev. Cult. Change* 59:187–229
- Pitt MM, Rosenzweig MR, Hassan MN. 2012. Human capital investment and the gender division of labor in a brawn-based economy. *Am. Econ. Rev.* 102:3531–60
- Prado EL, Sebayang SK, Apriatni M, Adawiyah SR, Hidayati N, et al. 2017. Maternal multiple micronutrient supplementation and other biomedical and socioenvironmental influences on children's cognition at age 9–12 years in Indonesia: follow-up of the SUMMIT randomised trial. *Lancet Glob. Health* 5:e217–28
- Price DJ, Song J. 2016. *The long-term effects of cash assistance*. Work. Pap., Stanford Univ., Stanford, CA
- Rahman MM, Vermund SH, Wahed MA, Fuchs GJ, Baqui AH, Alvarez JO. 2001. Simultaneous zinc and vitamin A supplementation in Bangladeshi children: randomised double blind controlled trial. *BMJ* 323:314–18
- Rahmathullah L, Underwood BA, Thulasiraj RD, Milton RC, Ramaswamy K, et al. 1990. Reduced mortality among children in southern India receiving a small weekly dose of vitamin A. *N. Engl. J. Med.* 323:929–35
- Ross D, Dollimore N, Smith P, Kirkwood B, Arthur P, et al. 1993. Vitamin A supplementation in northern Ghana: effects on clinic attendances, hospital admissions, and child mortality. *Lancet* 342:7–12
- Sazawal S, Dhingra U, Dhingra P, Hiremath G, Sarkar A, et al. 2010. Micronutrient fortified milk improves iron status, anemia and growth among children 1–4 years: a double masked, randomized, controlled trial. *PLOS ONE* 5:e12167
- Sempértegui F, Estrella B, Camaniero V, Betancourt V, Izurieta R, et al. 1999. The beneficial effects of weekly low-dose vitamin A supplementation on acute lower respiratory infections and diarrhea in Ecuadorian children. *Pediatrics* 104:e1
- Shankar AH, Genton B, Semba RD, Baisor M, Paino J, et al. 1999. Effect of vitamin A supplementation on morbidity due to *Plasmodium falciparum* in young children in Papua New Guinea: a randomised trial. *Lancet* 354:203–9
- Sivakumar B, Vijayaraghavan K, Vazir S, Balakrishna N, Shatrugna V, et al. 2006. Effect of micronutrient supplement on health and nutritional status of schoolchildren: study design. *Nutrition* 22:S1–7
- Skoufias E, McClafferty B. 2001. *Is PROGRESA working? Summary of the results of an evaluation by IFPRI*. Discuss Pap. 118, Food Consum. Nutr. Div., Int. Food Policy Res. Inst., Washington, DC
- Solon FS, Sarol JN Jr., Bernardo AB, Solon JAA, Mehansho H, et al. 2003. Effect of a multiple-micronutrient-fortified fruit powder beverage on the nutrition status, physical fitness, and cognitive performance of schoolchildren in the Philippines. *Food Nutr. Bull.* 24:S129–40
- Sommer A. 2008. Vitamin A deficiency and clinical disease: an historical overview. *J. Nutr.* 138:1835–39
- Sommer A, Djunaedi E, Loeden A, Tarwotjo I, West K Jr., et al. 1986. Impact of vitamin A supplementation on childhood mortality: a randomised controlled community trial. *Lancet* 327:1169–73
- Stansfield SK, Pierre-Louis M, Augustin A, Lerebours G. 1993. Vitamin A supplementation and increased prevalence of childhood diarrhoea and acute respiratory infections. *Lancet* 342:578–82
- Strauss J, Witoelar F, Sikoki B. 2016. *The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report*. Santa Monica, CA: RAND Corp.
- Suchdev PS, Ruth LJ, Woodruff BA, Mbakaya C, Mandava U, et al. 2012. Selling Sprinkles micronutrient powder reduces anemia, iron deficiency, and vitamin A deficiency in young children in western Kenya: a cluster-randomized controlled trial. *Am. J. Clin. Nutr.* 95:1223–30

- Suppl. Mult. Micronutr. Interv. Trial Study Group, Shankar AH, Jahari AB, Sebayang SK, Aditiawarman, et al. 2008. Effect of maternal multiple micronutrient supplementation on fetal loss and infant death in Indonesia: a double-blind cluster-randomised trial. *Lancet* 371:215–27
- Tanner JC, Candland T, Odden WS. 2015. *Later impacts of early childhood interventions: a systematic review*. Work. Pap., Ind. Eval. Group, World Bank, Washington, DC
- Thomas D, Frankenberg E, Smith JP. 2001. Lost but not forgotten: attrition and follow-up in the Indonesia Family Life Survey. *J. Hum. Resourc.* 36:556–92
- Thomas D, Witoelar F, Frankenberg E, Sikoki B, Strauss J, et al. 2012. Cutting the costs of attrition: results from the Indonesia Family Life Survey. *J. Dev. Econ.* 98:108–23
- UNICEF, World Health Organ., World Bank Group Joint Child Malnutr. Estim. 2018. *Levels and trends in child malnutrition*. Tech. Rep., UNICEF/World Health Organ./World Bank, New York/Geneva/Washington, DC
- Veenemans J, Milligan P, Prentice AM, Schouten LR, Inja N, et al. 2011. Effect of supplementation with zinc and other micronutrients on malaria in Tanzanian children: a randomised trial. *PLOS Med.* 8:e1001125
- Vijayaraghavan K, Radhaiah G, Prakasam BS, Sarma KR, Reddy V. 1990. Effect of massive dose vitamin A on morbidity and mortality in Indian children. *Lancet* 336:1342–45
- West KP Jr., Katz J, LeClerq SC, Pradhan E, Tielsch JM, et al. 1991. Efficacy of vitamin A in reducing preschool child mortality in Nepal. *Lancet* 338:67–71

