IDS WORKING PAPER
Volume **2011** Number **376**

Admissible Evidence in the Court of Development Evaluation?

The Impact of CARE's SHOUHARDO Project on Child Stunting in Bangladesh

Lisa C. Smith, Faheem Kahn, Timothy R. Frankenberger and Abdul Wadud
October 2011

Admissible Evidence in the Court of Development Evaluation? The Impact of CARE's SHOUHARDO Project on Child Stunting in Bangladesh

Lisa C. Smith, Faheem Khan, Timothy R. Frankenberger, Abdul Wadud

Summary

Along with the rise of the development effectiveness movement of the last few decades, experimental impact evaluation methods – randomised controlled trials and quasi-experimental techniques – have emerged as a dominant force. While the increased use of these methods has contributed to improved understanding of what works and whether specific projects have been successful, their 'gold standard' status threatens to exclude a large body of evidence from the development effectiveness dialogue.

In this paper we conduct an evaluation of the impact on child stunting of CARE's SHOUHARDO project in Bangladesh, the first large-scale project to use the rights-based, livelihoods approach to address malnutrition. In line with calls for a more balanced view of what constitutes rigor and scientific evidence, and for the use of more diversified and holistic methods in impact evaluations, we employ a mixed-methods approach. The results from multiple data sources and methods, including both non-experimental and quasi-experimental, are triangulated to arrive at the conclusions. We find that the project had an extraordinarily large impact on stunting among children 6–24 months old – on the order of a 4.5 percentage point reduction per year. We demonstrate that one reason the project reduced stunting by so much was because, consistent with the rights-based, livelihoods approach, it relied on both direct nutrition interventions and those that addressed underlying structural causes including poor sanitation, poverty, and deeply-entrenched inequalities in power between women and men. These findings have important policy implications given the slow progress in reducing malnutrition globally and that the widely-supported Scaling Up Nutrition initiative aimed at stepping up efforts to do so is in urgent need of guidance on how to integrate structural cause interventions with the direct nutrition interventions that are the initiative's main focus. The evaluation also adds to the evidence that targeting the poor, rather than employing universal coverage, can help to accelerate reductions in child malnutrition. The paper concludes that, given the valuable policy lessons generated, the experience of the SHOUHARDO project merits solid standing in the knowledge bank of development effectiveness. More broadly, it illustrates how rigorous and informative evaluation of complex, multi-intervention projects can be undertaken even in the absence of the randomisation, non-project control groups and/or panel data required by the experimental methods.

Lisa C. Smith is a senior economist at TANGO, International in Tucson, Arizona. She was formerly a Research Fellow at the International Food Policy Research Institute and Visiting Faculty at the University of Arizona and Emory University. As an American Association for the Advancement of Science Post-Doctoral Fellow, she was a food security advisor in USAID's policy bureau. Smith holds a Ph.D. in Agricultural and Applied Economics from the University of Wisconsin-Madison. Her areas of expertise are development economics, food and nutrition security, and gender.

Faheem Khan is currently the Chief of Party of CARE's SHOUHARDO II Program in Bangladesh. He holds a Masters in Information Engineering from City University, London and a Masters in Engineering with Business Management from the University of Brighton. He has 15 years experience in development, working in Bangladesh and Africa to bring tangible benefits to some of the poorest and most marginalised sections of society. His expertise lies in designing and managing complex development and relief programs, institutional strengthening, community mobilisation, and designing and implementing participatory monitoring and evaluation systems.

Timothy R. Frankenberger is the President and co-founder of TANGO International, a consulting firm in international development. He has over 30 years of experience in development activities. He was previously the Senior Food Security Advisor and Livelihood Security Coordinator at CARE, providing strategic technical support and guidance in food and livelihood security programming to 61 CARE country offices. Prior to this, he was a farming systems research specialist at the University of Arizona. He was also the founding editor of the Journal of Farming Systems Research-Extension. He did master's and doctoral work at the University of Kentucky in Anthropology.

AKM Abdul Wadud is head of the Monitoring and Evaluation unit of CARE's SHOUHARDO II program. He possesses over 20 years of diversified development experience, including management of field operations, training, and development and management of quantitative and qualitative monitoring and evaluation systems. His expertise lies in developing innovative systems to provide reliable and timely data and transforming this data to readily usable information for program management and external audiences. He is author of a number of publications of technical books and reports.

Contents

Figures and Tables

Acknowledgements

Introduction

Child malnutrition results in life-long, irreversible damage, compromising children's health, educational achievement, and productivity as adults (Victora *et al.* 2008). Success in attaining a wide variety of development goals, from eradicating poverty to increasing education levels and ensuring environmental sustainability, is dependent on improving children's nutritional status (SUN 2010). Yet progress in reducing malnutrition has been slow, especially so in the last few years of food-price and financial crises (World Bank 2010a). A consensus has thus recently emerged that reducing malnutrition must be a global priority area in which investment should be sharply stepped up (SUN 2010; Victora 2009; von Grebmer *et al.* 2010).

What types of interventions should be invested in to bring about major reductions in child malnutrition globally? There is now common agreement that a broad 'multi-sectoral' approach is needed encompassing two routes: (1) implementing 'direct nutrition' interventions, such as breastfeeding promotion; and (2) addressing deeper structural causes, for example poverty (UNICEF 2009; SUN 2010; Gillespie, Mason and Martorell 1996). Despite the acknowledgement that both are necessary, there has been a long-standing tendency to treat the two separately (Frankenberger 2001), with the direct interventions being seen as a 'short route' and structural ones a 'long route' (World Bank 2006). For example, a 2002 review of successful community nutrition programs concludes that: 'Nutrition program managers cannot normally influence contextual factors, at least in the short term' (Iannotti and Gillespie 2002: ix).

This view is reflected in the new Scaling Up Nutrition (SUN) Initiative (SUN 2010) that has emerged in response to the call for intensified global efforts to reduce malnutrition. Signed on to by a broad coalition of stakeholders – from governments and donor agencies to researchers and non-governmental organisations – the focus of the initiative is on 13 direct nutrition interventions, ranging from de-worming drugs to iron fortification of staple foods, that are considered to be the most high impact and cost effective. It is believed that these interventions, provided widely, will bring about 'rapid and immediate' reductions in child malnutrition at a minimum cost of $10 billion per year. While the SUN initiative recognises the need for taking a multi-sectoral approach, no specific plans regarding the implementation of interventions that address structural causes, for example, specifically which ones and how much they will cost to implement alongside the proposed direct nutrition interventions, are included.

In fact, the development community does not have a great deal of experience addressing malnutrition through interventions that address its structural causes. Consequently, there is a dearth of knowledge about the impact of these interventions[1] and on the possibly synergistic impacts of combining direct nutrition interventions with those that address structural causes.

In this paper we conduct an evaluation of the impact of the SHOUHARDO project[2] of CARE International, operating in Bangladesh from 2006–2010, on child malnutrition. This project stands apart from its predecessors in that it is the first large-scale development program that, in addition to direct nutrition interventions, including food aid, addressed a broad range of structural causes specific to the population in its area of operation using a rights-based, livelihoods approach (Frankenberger *et al.* 2000). These structural causes include not only poverty, poor sanitation, and recurrent natural disasters, but also deeply-entrenched inequalities in power between economic classes and between women and men. While,

---

[1]   Note, however, that there has been a great deal of research on the impacts of increasing income (e.g., Smith and Haddad 2000) and of sanitation improvements, which address the health environment (e.g. Esrey 1996).

[2]   The SHOUHARDO acronym stands for 'Strengthening Household Ability to Respond to Development Opportunities' and also means "friendship" or "amity".

traditionally, development agencies have shied away from these politically sensitive areas involving social inequalities, in project planning they were thought to be key leverage points for addressing the high prevalence of malnutrition among the project population in a sustainable manner.

The project also stands apart from its predecessors in the exceptionally large reduction in child malnutrition that took place over its implementation period. As illustrated in Figure 1, the prevalence of stunting, a measure of chronic malnutrition, among children 6–24 months old in the SHOUHARDO project's operational area declined from 56 to 40 per cent over the life of the project.[3] This drop of 4.5 percentage points per year is far higher than that in Bangladesh country-wide for this age group, which was only 0.1 percentage points per year in the first decade of the New Millennium.[4] It is also higher than the experience of previous nutrition-oriented projects. The average USAID-funded Title II food aid program is associated with a reduction in stunting prevalence among children under five years from baseline to final evaluation of 2.4 percentage points per year (Swindale *et al.* 2004).

Figure 1 Change in the prevalence of stunting among children 6–24 months old in the SHOUHARDO project's operational area, February 2006 – November 2009
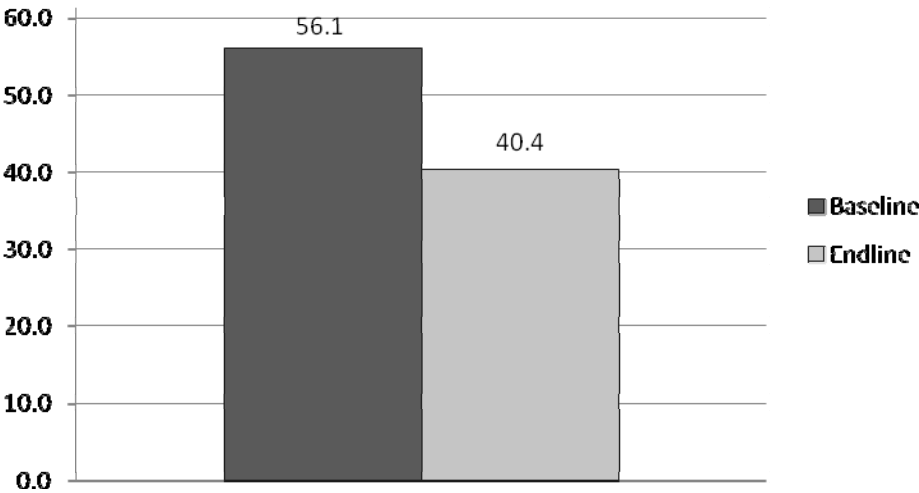
Notes: Prevalences are calculated using the data collected in the project baseline (February 2006) and endline (November 2009) surveys. They are based on the WHO 2006 growth standards.

Given limited knowledge about the integration of direct nutrition interventions with those that address deeper structural causes, if the SHOUHARDO project's interventions actually led to such unusually large declines in child malnutrition it is important to learn from and draw on its lessons. To do so, this paper investigates two key questions: First, was the observed reduction in stunting actually brought about by the SHOUHARDO project's interventions? Second, if so, *why*? Specifically, did the addition of a suite of structural cause interventions have added impact over the direct nutrition interventions? We also look into the role played by the careful targeting of the poor that took place, which is so fundamental to the livelihoods approach.

---

[3] Specifically, this decline took place between the time of the project's baseline and endline surveys, a 3.5 year period (see below for details on the surveys).

[4] The reduction (for rural areas only) was from 46.6 to 45.8 per cent between 2000 and 2010. The equivalent reduction for *all* under-fives is 0.56 percentage points per year (50.4 to 44.8 per cent). These statistics are derived from the authors' analysis of Bangladesh Demographic and Health Survey data collected in 2000, a Helen Keller International publication (HKI 2010) (for 0–59 month olds in 2010), and information provided directly by HKI staff (for 6–24 month olds in 2010). All prevalences are derived using WHO (2006) standards.

With the rise of the development effectiveness movement over the last decades there is now wide accord not only that more impact evaluations are needed but also, and most importantly, that more high quality evaluations based on rigorously-derived evidence be conducted (CDG 2006). In response to this demand, the experimental impact evaluation methods, that is randomised controlled trials (RCTs) and quasi-experimental techniques, have risen as a dominant force. These methods rely on random choice of project participants, on control groups, and/or on data collected from the same households over time, that is panel data, to precisely estimate the magnitude of impact of a project or intervention on intended outcomes. The methods are considered by some to be the 'gold standard' and, in fact, the only methods with sufficient rigor to be able to judge what works, whether implementation of specific projects has been successful and, in turn, which types of interventions and projects should receive future donor funding. This exclusive outlook threatens to bar a large body of evidence from the development effectiveness dialogue. And much is at stake, since information emanating from evaluations can ultimately have a powerful influence on the quality of life of developing country people.

As discussed in detail below, the methods of evaluation used in this analysis fall outside the bounds of experimental methods. Experimental techniques raise ethical issues in the context of food assistance and, further, have limited usefulness for evaluating complex, multi-intervention projects such as SHOUHARDO. The primary data used for the evaluation were collected from a random sample of participating households before the project began and near its end. This design does not allow precise estimation of the magnitude of the project's impact. We demonstrate here, however, how a judicious and intelligent use of the available data, combined with key information about changes in the project's external environment over its implementation period and from project administrators, can provide rigorous, informative and useful evidence regarding its impact on child stunting.

The next section gives an overview of the SHOUHARDO project. Following, an account of the evolution of conventional methods of quantitative impact evaluation is given and the methods used in this analysis are laid out. The evidence on whether the SHOUHARDO project activities led to the observed decline in stunting is then presented. Further analysis is undertaken that both provides complementary evidence of the project's impact on stunting and illuminates why the project had this impact. The paper concludes with a summary of the empirical findings and some reflections on admissible evidence in the court of development evaluation.

1 The SHOUHARDO Project

The SHOUHARDO project was one of the largest in the world in the portfolio of the United States Public Law 480 Food for Peace Title II food aid program at the time of its implementation. Having primary goals to reduce child malnutrition, poverty and food insecurity, and serving a population of two million people, the program was carefully targeted to the most remote and vulnerable areas of the country and, within these areas, to the poorest households. Here, we describe this targeting process, the rights-based, livelihoods approach used for project planning, and the plethora of interventions that were implemented.

Targeting of participant households

A systematic targeting process, including both geographical and household targeting, was employed to identify the poorest households in Bangladesh to be SHOUHARDO project participants. In a first step, national databases were used to identify the remote areas most vulnerable to shocks and food insecurity. The four regions found were populated by some of the most marginalised groups in the country due to their uniquely adverse agro-climatic

conditions that inhibit food production and economic activity. The *North Char* and *Mid Char* regions are situated in areas of unstable land that appears and disappears with accretion and erosion of sandy riverbed soils, leading to periodic flooding and consequent river erosion. The *Haor* region, with its tectonically depressed areas comprised of a river network fed by rainfall in the Indian hills is submerged from May to October. During the monsoon season human settlements are confined to densely populated, isolated man-made mounds. In the *Coast* region, silt-poor soils, tidal waves and seasonal storm surges are chronic limitations to agricultural productivity, and tropical storms are constant threats. After identifying these areas, CARE staff carried out extensive consultations with government officers and civil society organisations in each region to select the poorest villages in them for inclusion in the project's operational area.[5]

Household targeting within each village began with a participatory 'well being analysis'. Community members representing the broad range of interest groups and classes grouped households into four economic categories: extreme poor, poor, middle class and rich. The classification criteria used included land ownership, housing condition, income level, income sources, occupation and food insecurity. The extreme poor and poor were chosen as project participants after CARE staff physically visited each household to confirm eligibility. The finalised participant list included 400,000 households representing on average three-quarters of all households in project villages.

The rights-based, livelihoods planning approach
As mentioned above, project planners chose interventions using a rights-based, livelihoods approach. The *livelihoods approach*, which came to fore in the mid 1990s, takes a holistic perspective on people's lives to inform programming decisions. It is based on the principle that, no matter what outcome a project is intending to influence, it is important to have a comprehensive understanding of the strategies that households use to negotiate survival. This requires taking into account the trade-offs faced in all major areas of their lives, including food security, health, education, income, and physical security. The approach leads to a breakdown of the traditional sectoral view in which people's lives are fragmented into unconnected pieces. It thus helps clearly identify opportunities and leverage points for positive change. Inherent in the approach is a search for cross-sectoral combinations of interventions that have synergistic impacts (Frankenberger *et al.* 2000; Drinkwater 2001; Jennings and McCaston 2007).

The *rights-based approach*, given global legitimacy by the Universal Declaration of Human Rights, challenges development practitioners to take responsibility for dealing with issues of inequality, discrimination and rights denial. It is founded on the notion that claims on rights are critical to households' livelihoods (Drinkwater 2001; Frankenberger 2001).

Writes Drinkwater (2001: 3), 'Combining rights and livelihoods approaches ensures a grounding of all forms of development work in a respect for human dignity, in principles of social justice, in the need to address power relations and the causes of poverty, vulnerability and marginalisation more seriously'. The combination also enhances understanding of social sustainability and, subsequently, provides a basis for more sustainable development (Moser and Norton 2001).

Project interventions
The SHOUHARDO project's MCHN interventions were employed to address the most immediate causes of child malnutrition using many of the direct nutrition interventions recommended by the SUN initiative. The rights-based, livelihoods planning approach revealed five key structural constraints to reducing malnutrition in the project area: unsanitary

---

[5] A small number of urban slums (containing 3 per cent of project households) with quite different environmental and economic conditions than the project's rural majority were included. This paper's analysis focuses only on the rural areas.

living conditions, discrimination against women, poverty, discrimination against the poor, and recurrent natural disasters. Here, starting with MCHN, the interventions that were designed to address these constraints are described.

Mother and child health and nutrition (MCHN). The project's most direct nutrition intervention was the provision of food rations to children from 6 months to 2 years of age and to pregnant and lactating mothers. Children were eligible to receive food equivalent to 250 kilocalories per day, their mothers 1,900 kilocalories per day. Health and nutrition education was provided through 'mother's' group sessions, with topics covering optimal breastfeeding, complementary feeding and weaning practices, care for mothers during pregnancy and delivery, and hygiene practices. Additional interventions were child growth monitoring, provision of pre-natal care, birth planning, and aid in obtaining emergency obstetric care for mothers; Vitamin A supplementation for children and Vitamin A and iron-folic acid supplementation for mothers; immunisations; referrals for family planning and emergencies; and facilitation of access to local health facilities. These services were provided by community health volunteers who were also responsible for convening the mother's group sessions and monitoring compliance with the recommended health and child feeding practices. In line with the preventative approach to reducing child malnutrition (Ruel *et al.* 2008),[6] all MCHN interventions were targeted to children under two and their mothers.

Sanitation. Diarrheal disease is a key cause of child malnutrition in Bangladesh, with lack of access to safe water and sanitary latrines being its main structural cause (IRIN 2010). The SHOUHARDO area had a particularly high prevalence of diarrhea at the start of the project, at 22 per cent. While 70 per cent of households had access to safe water, only 22 per cent had access to a sanitary latrine. The project addressed this problem by assisting households in obtaining safe, arsenic-free drinking water through the installation of tubewells and arsenic testing, as well as access to sanitary latrines.

Empowerment of women. Discrimination against women is strong and pervasive in Bangladesh; women's status is among the lowest in the world (Rozario 2004; Smith *et al.* 2003). Enhancing women's empowerment was considered by project planners to be crucial for improving children's nutritional status because of its important role in enabling women to receive proper care for themselves as childbearers and provide adequate care for their children (Smith *et al.* 2003). The central intervention designed to do so was Empowerment, Knowledge and Transformative Action (EKATA) groups, which established a recognised and accepted forum for women to meet and express themselves in a public role. The groups, comprised of twenty women and ten adolescent girls, provided a platform for empowering women and girls through education, solidarity, group planning, and rights advocacy. Another intervention was the introduction of Early Child Care for Development (ECCD), a preschool that introduces a learning process, flow of information, and preparation for formal schooling that had been traditionally denied to girls. The project also promoted Parent Teacher Associations (PTAs)[7] to facilitate participation of women in the formal educational structure and the education of their daughters. Together, the three interventions had a broad range of goals: increasing women's decision making power at household and community levels, reducing gender-based violence, raising awareness of educational entitlements for women and girls, building women's leadership, advocacy, and literacy skills, and consciousness-building around important social issues, including dowry, early marriage, divorce, and violence against women.

Poverty and food insecurity alleviation. A suite of interventions was introduced to address the extreme poverty and food insecurity in the project area. To increase food production and incomes, training and inputs were provided to promote field crop and fisheries production,

---

[6]   A description of the preventative approach can be found in the introduction to Section 5 below.

[7]   Another group, a School Management Committee, was also established. Herein this dual intervention is simply referred to as Parent Teacher Associations.

homestead gardening and livestock rearing, and cash income generating activities. Economic support was also provided through food-for-work and cash-for-work, which not only gave temporary employment but also resulted in infrastructural improvements from which entire communities benefited, such as roads, protective walls, and market centers. Finally, the project facilitated the establishment of savings groups to help households pool financial resources for emergencies, acquire assets, and leverage new business opportunities.

Empowerment of the poor. Underlying poverty in Bangladesh is the unequal and exploitative structure of power in local society. For centuries local elites have controlled access to resources through land ownership and control of water bodies, influence over the justice system, preferential political access, and even thuggery. To create the conditions for a transformation of this traditional power structure, a collective institution coalescing the power of and providing representation for the poor, the 'Village Development Committee' (VDC), was established in all project villages. Committee members were elected by a broad cross-section of community members and authorised to create and operationalise a community-driven development plan. The VDCs were then integrated into the formal structure of government from the local union up to national levels. In addition to collaborating with CARE staff to implement project activities and the community development plan, the VDCs were responsible for managing local resources for the common good and for defending the interests of the poor. They promoted rights awareness among poor households and advocated on their behalf to secure rights to common land resources and prevent eviction.

Disaster mitigation and response. Finally, the project addressed a major long-standing obstacle to all aspects of development among the very poor in Bangladesh, the persistent threat of natural disasters due to flooding and cyclones that prevents households from making any long-term investment in their future livelihoods. Activities helped to develop local institutional capacity to prepare for and respond to disasters, build capacity for quick response, and develop infrastructure for disaster mitigation.

2 The evolution of conventional methods of quantitative impact evaluation

How does one go about evaluating whether and how such a complex, multiple-intervention project as SHOUHARDO was able to bring about a reduction in stunting in its area of operation?

In its broadest sense, impact evaluation is about determining the extent to which changes in outcomes can be *attributed* to a project or intervention. The 'counterfactual' model of impact evaluation provides a strong conceptual foundation for doing so. Here, evaluating attribution requires comparing what happened to the outcome *with* an intervention (the factual) to what would have happened to the outcome *without* it (the counterfactual). The counterfactual is never known with certainty because the participants in an intervention cannot **not** participate in it at the same time. Thus the evaluator must construct a credible substitute counterfactual in the form of a non-participant control group. However, households that do participate in an intervention are typically different from those who do not, whether due to targeting by project administrators or systematic differences among those choosing to participate and those not. A central challenge of impact evaluation is to eliminate the resulting selection bias by correcting for these differences (Ravallion 2001; Leeuw and Vaessen 2009; Khandker, Koolwal and Samad 2010).

Historically, few (among the few) quantitative impact evaluations of international development activities were founded on these basic principles (White and Bamberger 2008). More recently, an approach that has emerged to step up to the challenge of dealing head-on with selection bias is the use of experimental methods, that is, randomised controlled trials (RCTs) and quasi-experimental methods. An RCT attempts to achieve unbiased assessment of impact through randomly assigning households to be either participants or non-participants (controls) in an intervention before it is implemented. Because members of these groups do not differ systematically before implementation, any difference that subsequently arises between them can be attributed to the intervention and not to other factors (J-PAL 2011a). Quasi experimental methods are used to control for factors leading to selection bias using statistical techniques. One such method believed to be most effective at doing so is difference-in-difference propensity score matching (PSM), in which data from a panel survey[8] of participants and non-participants are first used to identify a comparable control group. Following, the change over the intervention implementation period for participant and control group households in the outcome of interest is used to estimate impact (Khandker, Koolwal and Samad 2010).

The wide range of impact evaluation methods in use extend from such experimental methods to qualitative techniques based on peoples' own views of how development activities have impacted their lives. A main advantage of experimental techniques over others is that, if properly implemented, they are able to precisely estimate the *magnitude* of impact of a project or intervention on intended outcomes.[9] Some go so far as to define impact evaluation itself based on this ability. According to researchers at the Abdul Latif Jameel Policy Action Laboratory (or 'J-PAL'), which relies exclusively on RCTs for evaluation, 'The primary purpose of impact evaluation is to determine whether a program has an impact… and more specifically, to quantify *how large* that impact is' (J-PAL 2011b).[10]

Reflecting the dominance and rising influence of the precisionist evaluation culture in international development impact evaluation over the last two decades, experimental evaluation techniques are widely believed to be at the top of the hierarchy of rigor, with RCTSs considered by many to be the 'gold standard' (Jones N. *et al.* 2009; J-PAL 2011). Some have even argued that projects should not be funded if they have not been shown to have positive impacts in an RCT (Banerjee 2007). However, with the experience of time, professional evaluators and development researchers have begun to question and critically evaluate the universal validity and – given their narrow focus – usefulness, of evaluations that rely exclusively on randomised designs (EES 2007; Ravallion 2009; Deaton 2010; Barrett and Carter 2010).

More generally, rigid adherence to one dominant evaluation paradigm, with its accompanying 'delegitimization of other ways of knowing' (Barrett and Carter 2010: 527), is increasingly being viewed as a setback to development progress. It can skew research and the dialogue on development effectiveness towards interventions that are evaluable using experimental techniques, with the ultimate implication that methodology becomes a driver of intervention choice and resource allocation (Jones 2009; Ravallion 2009, 2011; Barrett and Carter 2010). Further, it is thought to discourage risk taking and innovation, especially for projects addressing complex and long intractable problems, projects that may not be amenable to evaluation using these techniques (Jones H. 2009; Lawrey 2010). Taken to an extreme, it

---

[8]   A panel survey is one where the same households or individuals are included in the sample at different points in time.

[9]   Another putative advantage of RCTs is that relatively few assumptions are required to establish that a causal impact has been identified, making results easier to communicate to policy makers (Duflo and Kremer 2003; J-Pal 2011c). Barrett and Carter (2010) give a counter view.

[10]   Similarly, the International Initiative for Impact Evaluations (or "3IE') defines rigorous impact evaluations as 'analyses that measure the net change in outcomes for a particular group of people that can be attributed to a specific program using the best methodology available, feasible and appropriate to the evaluation question that is being investigated and to the specific context' (IIIE 2011a: 1).

can lead to modifications of project design and implementation that may be detrimental to project success.

The selective filtering out of evaluations that are not based on the dominant techniques inhibits the learning that is so crucial to the success of the development effectiveness agenda. This filtering takes many forms, including: donors being unwilling to fund non-experimental evaluations; researchers turning down evaluations when experimental designs are not feasible; Doctoral students searching out research topics where randomisation is possible (Ravallion 2009, 2011); development decision makers being told that the questions they ask that can't be answered using experimental methods are the 'wrong questions' (Barrett and Carter 2010: 528); large bodies of useful, and sometimes essential, prior knowledge based on non-experimental techniques being ignored in experimental evaluations (Barrett and Carter 2010); and articles and reports of evaluations being excluded from impact evaluation databases (see IIIE 2011a,b) and rejected from professional development journals because they do not use experimental methods.

An estimate of the exact magnitude of impact of a project's intervention(s) on intended outcomes is valuable information for determining which interventions work, whether implementation of specific projects has been successful and, in turn, which interventions and projects should receive donor funding. However, they are not necessary for answering these questions, and not always appropriate.

Given the availability of a wide variety of alternative methods, evaluators are increasingly being challenged to open their minds to a more balanced view of what constitutes rigor and scientific evidence (Jones H 2009; Rugh 2011). The pendulum has now started to swing towards a more methodologically diverse and eclectic approach to impact evaluation while still ensuring acceptable standards of rigor (ESS 2007; Barrett and Carter 2010; Ravallion 2011; Rugh 2011). There is a growing demand for alternative approaches for assessing counterfactuals and, more generally, for ways to conduct evaluations that give proof of impact while not being subject to the precision of an RCT or quasi-experimental methods (Rugh *et al.* 2010). One element of this paradigm shift is a growing consensus around the use of a more holistic, mixed-methods approach to strengthen the quality of evaluations. Such an approach can be used to assess different facets of impacts, yielding a broader and richer portrait than one method alone. Further, triangulation of results from a variety of methods, including qualitative methods, can increase validity and confidence in the findings of an impact evaluation. If the results of different methods converge, then inferences about impact are stronger (White 2008; Leeuw and Vaessen 2009; Khagram *et al.* 2009; Chambers 2009; Rugh 2011).[11]

The practical importance of a more diverse approach to impact evaluation for the success of the development effectiveness mission becomes clear as one considers that the large majority of development activities do not incorporate an experimental evaluation design – for a long list of reasons. These include: ethical concerns, lack of resources, political concerns, unwillingness to alter project design to accommodate the methods, the impossibility of identifying a control group, an inappropriate match of method with the type of intervention (e.g., macroeconomic interventions), and the difficulty of measuring key outcomes using standardised quantitative indicators (Duflo and Kremer 2003; Deaton 2007; Ravallion 2009; Rugh 2011; Roetman 2011). Rugh (2011) estimates that experimental methods are either not possible or inappropriate in 95 per cent of impact evaluations.

---

[11]   Indeed Khagram *et al.* (2009) boldly write that 'The platinum standard of rigour for impact evaluations (IEs) to generate credible evidence is the systematic inclusion and application of multiple types of comparison and triangulation' (p. 248).

3 Methods and data used for this evaluation

Overall approach

The SHOUHARDO project is typical of the large majority of projects for which an experimental or quasi-experimental design cannot be used for the evaluation of its impact. The principal reason is the lack of a comparable, non-project control group to serve as the counterfactual.

As is common for projects with nutrition goals, especially those distributing food to highly food insecure populations, it was not considered ethical to randomise participation, thereby excluding some eligible households for the purposes of evaluating the project (Levinson *et al.* 1999). Furthermore, doing so would have violated one of the basic principles on which the project was founded, which was that the project's scarce resources should be targeted to the poorest households in the country.[12]

Furthermore, selection of a control group from a sample of non-project households, whether using a random design or not, was not possible. Most of the population of interest – that is poor households living in remote and areas vulnerable to shocks – was already located within the project area,[13] ruling out the location of a comparable control group affected by the same environmental conditions. Although it may have been feasible to identify a control group of households from a sample of non-project households using statistical matching methods such as propensity score matching, because the targeting for the project was so precise, the required sample size may have been quite large (and prohibitively costly) to be able to locate sufficient matches.

Quite apart from this sample size issue is the fact that, given the large number of outcomes related to child malnutrition the SHOUHARDO project was attempting to influence (from health behaviors to food security and incomes), there were few important characteristics left on which matching could be made. A national household survey conducted by Helen Keller International (HKI) in which data were collected on children's nutritional status was conducted at roughly the same time as the project's endline survey (see below for details). An attempt was made to use these data to identify a comparable control group as done, for example, by Handa and Maluccio (2010) and Donnegan *et al.* (2010). It was found that this was not possible because virtually all of the indicators for which data are available in both the project (endline) and HKI data sets with which matches could be made were intended to be affected by the project.[14] If matches are made on these indicators, then project households are matched not with those that would provide an appropriate counterfactual, but simply with households that are already as well off, for whatever reason, as project households after they had received the interventions. Other projects with a wide range of interventions and intended outcomes are likely to run into the same obstacle.

Another factor complicating the identification of a comparable, 'uncontaminated' control group is the fact that many of the project's interventions likely had spillover effects, especially the disaster mitigation and response and empowerment interventions. For example, one of the main goals of the establishment of the VDCs was to increase citizen responsiveness of government bodies in the wider region not just in the local area of the villages.

---

[12] Barrett and Carter (2010) give an thorough discussion of the ethical issues involved in randomised experimental evaluations, which in many cases apply to the use of control groups in general.

[13] There are two exceptions. The first is poor households in the south west coastal belt of the country, who were excluded because other international NGOs already had large-scale food security and nutrition oriented projects there. The second is poor households in three (out of 64) of the country's districts that were excluded because CARE did not have the organisational capacity to set up operations in them.

[14] Only demographic indicators, such as households' age and sex composition and location were available for performing matching.

As noted in the introduction, the primary survey data available for this analysis are cross-sectional baseline and endline surveys of project participants. These kinds of data do not allow the use of experimental methods to precisely estimate the magnitude of the overall impact of the SHOUHARDO project on child stunting. To determine whether there is plausible evidence that the measured decline over time can be attributed to the project's interventions and, if so, why, we employ the mixed methods approach being advocated by proponents of more holistic impact evaluation, relying on a variety of primary and secondary data sources and statistical methods. This approach is seen to be particularly important in the case of complex, multi-intervention projects like SHOUHARDO, for which more specific methodological clarification is not available but sorely needed (Elbers, Gunning and Hoop 2009; Rugh 2011; Ravallion 2011).

For evaluating the overall impact of the project we use the available baseline and endline survey data, data from other national surveys, and information about what was going on in Bangladesh as a whole over the project implementation period. For investigating the 'why' question, we make use of PSM in conjunction with supplemental information collected from project administrators on the selection of households for participation in specific project interventions, as well as descriptive analysis of heterogeneous impacts across sub-groups of project households. The results of all analyses are triangulated to reach the study's final conclusion regarding the impact of the project on stunting.

Data and measurement of nutritional status

The SHOUHARDO project baseline survey (N=3,200) was conducted in February 2006. The survey was administered to households with children 6–24 months old, the target group for MCHN interventions. Two endline surveys were conducted. The first (N=3,200), conducted in August 2009, was administered to households with children 48–59 months old in the same villages as the baseline to help investigate whether the project had long-term nutritional benefits. The surveyed children would have been 6–18 months at the time of the baseline survey. Only nutritional data were collected. The second endline survey (N=3,356) was conducted in November 2009 from a newly-drawn random sample of project households with children 6–24 months old. In addition to nutritional data, data were collected on a large number of household characteristics and outcomes as well as on the interventions in which they participated. All surveys were conducted using a two-stage, stratified random sampling design, where the four project areas were the strata and villages the primary sampling units.[15]

Data collected in the surveys on each child's height and age were used to calculate the two nutritional outcomes focused on in this paper: height-for-age z-score and stunting. Stunting is a result of inadequate growth of the fetus and child and results in a failure to achieve expected height compared to a healthy, well-nourished child of the same age. It is an indicator of past growth failure and is associated with long-term factors including chronic insufficient protein and energy intake, frequent infection and sustained inappropriate feeding practices. It is calculated by combining the height and age data to calculate a height-for-age z-score. If the z-score is less than -2 standard deviations below the median of an adequately nourished reference population, the child is considered to be stunted (Cogill 2003). The reference population employed for this analysis is that used to develop the World Health Organization 2006 Child Growth Standards (de Onis *et al.* 2004).

In addition to the primary survey data collected, this evaluation relies on various sources of secondary data. The first is nationally-representative household surveys conducted since the

---

[15]   The baseline survey was conducted by The Asia Foundation. Extensive training in the collection of anthropometric data was provided by the International Centre for Diarrheal Disease Research, Bangladesh (ICDDR,B). The endline survey was conducted by the ICDDR,B with technical input and oversight from Technical Assistance to Non-Governmental Organisations, International (TANGO).

year 2000 in which data were collected on children's nutritional status (described in Section 4.1 below). The second is publications documenting the key factors affecting food and nutrition security countrywide in Bangladesh over the project evaluation period. The third is qualitative data collected from project administrators on the criteria used for targeting project areas and households for various interventions as well as the degree to which households self-selected into interventions.

Methods for evaluating the project's impact on stunting
To establish whether the project had an impact on child stunting – and the likely magnitude of that impact – three analyses are undertaken. The first is a comparison of the change in stunting in the project area with the change nationally over the same time period. This 'difference-in-difference' analysis is important for ruling out the possibility that the change in stunting prevalence was due to forces external to the project, for example a general upturn in the economy. Specifically, we examine the change in stunting among children 6–24 months in the project area between the baseline and endline surveys compared to the change that took place for this age group in rural areas countrywide (which serves as the counterfactual) using data from several national surveys. To interpret the differences, we bring together key information about what was happening in Bangladesh as a whole over the project period, which included a major food price crisis, unusually strong flooding, and a cyclone. Note that the project's population of two million people is very small relative to that of Bangladesh as a whole (roughly 150 million) such that changes in the project area had negligible influence on the stunting prevalence country wide. As a cross-check on the findings, we look at (1) changes over time in stunting for the four project areas, which are widely dispersed across the country; (2) the change over time in the project area compared to that in nearby non-project areas and for the poorest quintile of Bangladeshi households; and (3) evidence of the presence of non-SHOUHARDO project development activities in the project area.

The second analysis is a comparison of the age trajectory of the stunting prevalence among project children compared to children living in rural households nationally. The August 2009 endline survey of the age cohort of children who were 6–18 months old at baseline and 48–60 months old at endline is used to conduct this analysis. As will be shown, stunting typically shows a large increase over these age groups. We explore whether the change for the children who were exposed to project interventions shows a different pattern and discuss the implications for project impact.

In the third analysis we examine trends in the underlying determinants of child malnutrition – food security, quality of caring practices for children and women, and household health environment quality (UNICEF 1998) – over the life of the project. Because improvements in these determinants are necessary for bringing about reductions in child malnutrition, if such improvements indeed exist, they would provide further supporting evidence that malnutrition declined in the project area.

Methods for investigating why the project impacted stunting
A key piece of information that is needed to understand the 'why' question is whether the fact that the project addressed structural causes of malnutrition contributed to the reduction in stunting. Was the reduction due solely to the MCHN 'direct nutrition' interventions, including the monthly distributions of food aid, or did the interventions that addressed deeper causes and were likely to set in motion sustainable impacts, contribute as well? Did the MCHN and these other interventions have the synergistic impacts anticipated by the livelihoods approach?

Drawing on the considerable heterogeneity in participation of project households in individual interventions, we investigate these questions using PSM to create comparable-on-observables control groups for each intervention from among households that did not participate to serve as the counterfactual. The data from the November 2009 endline survey

are employed for this analysis.[16] To isolate the independent impact of each intervention, the fact that there may be differences in participation in the *other* project interventions across the participant and control groups is controlled for in the analysis.

There are two interventions that cannot be evaluated using PSM in this manner: disaster mitigation and response and empowerment of the poor (VDCs). This is because the interventions were implemented at the village or regional, not household, levels. Note that there is great variation in the degree to which households participated in the various MCHN interventions, with only 33 per cent *fully* participating. A household is considered to have fully participated in the MCHN intervention set if it received food aid, participated in growth monitoring, and participated in health and nutrition education.[17] We use this variation to help determine the direction of impact of the MCHN interventions as a whole.

The matching process in PSM takes place using measured indicators of characteristics that are believed to influence participation in an intervention as well as those influencing the outcome of interest, in this case nutritional status. If these *observed* characteristics are the only ones influencing participation, the estimates are deemed unbiased and the important 'conditional independence' condition is met. However, if *unobserved* characteristics also influence participation, then the estimates will be biased (Khandker, Koolwal and Samad 2010). Because panel data are not available to help account for any unobserved characteristics, we pay close attention to the likely direction of selection bias when interpreting the PSM results.

For any intervention, the PSM estimates of impact are generated in three steps. The first is to estimate a probit[18] participation model using data on both participants and non-participants to compute a probability of participation, or 'propensity score', for each household conditional on the observed characteristics. In the second step, participant households are matched with non-participant households based on similarity of propensity scores. An important condition for the success of this step is 'common support'. Participant households must be similar enough to non-participant households in the observed characteristics so that there are sufficient non-participant households close by in the propensity score distribution with which to make matches (Khandker, Koolwal and Samad 2010). For each intervention, common support is verified through examination of the propensity score distributions of participants and nonparticipants. Participant propensity scores that are higher than the maximum or lower than the minimum of the non-participant propensity score distribution are dropped.[19] In the third step of PSM, the average HAZ of the matched participant and non-participant groups of households are compared to calculate an estimate of the impact of the intervention, or the 'average treatment effect on the treated' (ATT).

To test for synergies[20] between MCHN and the other interventions we employ multiple treatment PSM (MT-PSM) (Imbens 2001; Lechner 2001).[21] The goal is to find out whether combining MCHN and sanitation interventions, for example, results in greater impact than could be achieved by introducing either intervention set alone. To do so, a reference case of no participation in either intervention is compared with three alternatives: (1) participation in

---

[16]   See Rawat, Kadiyala and McNamara (2010) for an example of another study that relies on sub-groups of project participants for intervention-specific control groups to evaluate a multi-intervention project.

[17]   A household is considered to have participated in growth monitoring if at least 90 per cent of the expected growth monitoring sessions given the child's age were attended.

[18]   A probit model is a dichotomous dependent variable regression model, that is, one where the dependent variable has two values (typically 0 and 1).

[19]   More stringent conditions were experimented with by trimming the participant distribution at the extremes. Doing so made minimal difference to the ATT estimates.

[20]   'Synergy' is defined, in its broad sense, as 'The action of two or more substances, organs, or organisms to achieve an effect of which each is individually incapable.' (American Heritage Dictionary, 2009). Corning (2007), of the Institute for the Study of Complex Systems, defines it as 'Otherwise unattainable combined effects that are produced by two or more elements, parts or individuals' (p. 1).

[21]   Examples of applications of the multiple treatment model, which are rare, can be found in Blundell, Dearden and Sianesi (2004), Becker and Egger (2007), and Moreno-Serra (2009).

MCHN only; (2) participation in the other intervention only; and (3) participation in both. Propensity scores are calculated using multinomial logit, which takes into account the simultaneity of the participation decision across multiple interventions. After matching, the mean HAZ of the groups of households falling into the three alternatives are compared with the no participation group to test for impact synergies.

Of the many techniques available, we conduct PSM and MT-PSM using kernel matching, for which each treated household is matched to a group of non-treated households with propensity scores within a certain radius.[22] The control group outcome is computed as a weighted average, with a lower weight given the greater is the propensity score difference from the treated household. The analysis is conducted using PSMATCH2 in STATA along with PSTEST to test for matching effectiveness (Leuven and Sianesi 2003). Matching effectiveness is evaluated by conducting t-tests for equality in the mean values of the characteristics on which matching is based across the participant and matched non-participant groups of households. An overall summary measure is given by the p-value from a likelihood ratio test for the joint insignificance of the characteristics *after* matching (that is, using the matched sample only). If the characteristics are no longer jointly significant ($p>0.10$), then matching has succeeded.

A second 'why' area addressed is whether targeting the poorest households rather than extending project interventions to all households in the project's operational area contributed to the reduction in stunting. To investigate this, we simply compare the total reduction in stunting over the life of the project for two groups of households, the 'extreme poor' and the 'poor', as identified during the initial targeting process (see Section 1). If the reduction is higher for the extreme poor, then we can plausibly conclude that the reduction is greater than it would have been if all households, including the 'middle class' and 'rich' ones, in the project's geographical area of operation where included as participants.

4 Evidence that the project's activities led to a large decline in stunting

4.1 Trends in stunting in project area compared to nationally

Figure 2 reports on the change in the prevalence of stunting among 6–24 month olds in the SHOUHARDO project's operational area between the project baseline and endline surveys compared to trends in rural Bangladesh over the same period. The surveys used to track national trends are Bangladesh Demographic and Health Surveys (BDHS) conducted in 2004 and 2007, surveys conducted by the National Nutrition Surveillance Project of Helen Keller International (HKI) in 2006 and 2010, and the Bangladesh Household Food Security and Nutrition Assessment conducted by the World Food Program, UNICEF, and Mitra and Associates from November 2008 to January 2009.[23] Note that the national surveys undertaken closest to the beginning and end of the project evaluation period were conducted quite close in time to them. The 2006 HKI survey and SHOUHARDO project baseline surveys were both conducted in February 2006. The SHOUHARDO endline survey was conducted in November 2009, and the comparison HKI survey was conducted shortly after, from January to April 2010.

---

[22] The radius depends on the bandwidth of the kernel, the choice of which involves a trade off between bias and variance, with smaller bandwidths increasing variance but leading to less biased estimates (Caliendo and Kopeinig 2008). After finding that variations between 0.01 and 0.1 make little difference to the ATT estimates, a bandwidth of 0.05 is used for all estimates.

[23] For more information about the data collection see NIPORT *et al.* (2005, 2009), HKI (2010) and WFP, UNICEF and IPHN (2009).

Figure 2 Change in stunting prevalence among children 6–24 months: project area versus nationwide (rural)
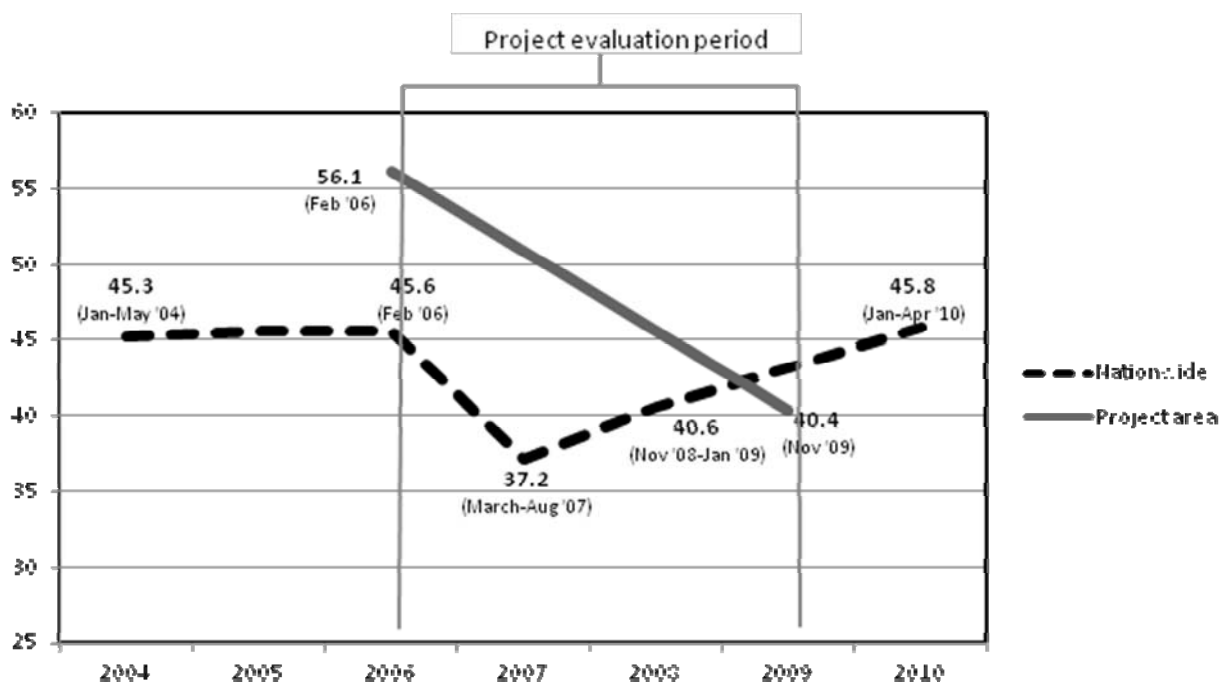
Sources: Project prevalences: From the baseline and November 2009 endline surveys. Nationwide prevalences: The five data points on which the trends are based are, in temporal order, from analysis of the data in the following surveys: Bangladesh Demographic and Health Survey (BDHS) 2004, National Nutrition Surveillance Project February 2006 survey conducted by Helen Keller International (NNSP-HKI), BDHS 2007, the Bangladesh Household Food Security and Nutrition Assessment conducted by the World Food Programme, UNICEF and the Institute of Public Health Nutrition (2008/09), and the NNSP-HKI survey conducted in January-April 2010.

SHOUHARDO project households saw a rapid reduction in the prevalence of stunting of 15.7 percentage points over the evaluation period. In sharp contrast, there was *no* decline in stunting in rural Bangladesh as a whole over this period and in fact an increasing trend was apparent in the project's last two years. In February 2006 the national prevalence of stunting for 6–24 month olds was 45.6 per cent. The prevalence declined by 2007,[24] at which point it began to show an increasing trend. It had risen to 45.8 per cent by the time of the project's endline survey, roughly the same prevalence as at the time of the baseline.[25] Note that while the prevalence of stunting among children of the SHOUHARDO population was far higher than that nationally at the start of the evaluation period (10.5 percentage points), by the end it was about five percentage points lower.

What explains the country-wide increase in stunting prevalence in the latter years of the project's implementation? In line with declining trends since the early 1990s (HKI 2006), the prevalence of stunting in rural Bangladesh fell by roughly 1.3 percentage points a year in the first seven years of the new Millenium. The subsequent abnormal increase in stunting prevalence took place as a result of a powerful food price shock that occurred in 2007 and

---

[24]   While seasonality affects stunting prevalences less than underweight and wasting prevalences, this drop is partially due to the fact that the 2006 HKI survey was conducted in February, which is a favorable time of year for food security, and the 2007 BDHS survey was conducted from March to August, the latter three months of which fall into the country's lean period (see HKI 2006).

[25]   Examination of trends by month of age for children under five shows clearly that it was mainly among children under two that the increase took place between 2007 (DHS data) and 2010 (HKI data). Since a similar increase did not take place for the older age groups such a sharp increase is not found for the entire under five group, the age group for which statistics on national stunting rates are normally reported (HKI 2011).

2008, mid-way through the SHOUHARDO project cycle. The shock was caused by rising global food and fuel prices associated with the global food price crisis. Adverse weather conditions that lead to rice crop losses contributed as well, including unusually high monsoon rains from Nepal and India in August-September 2007 led to the third largest floods in more than 50 years and Cyclone Sidr that struck coastal Bangladesh in November 2007.[26] During the shock, the prices of the main staple rice, and of pulses and edible oil, nearly doubled in Bangladesh. Analysis of data collected in a national survey conducted to investigate the impacts of the crisis show that households' purchasing power, real incomes, and food security deteriorated considerably over the period of the crisis (WFP, UNICEF and IPHN 2009; World Bank 2010a).

The evidence on comparative trends presented here is strongly suggestive that the decline in stunting seen in the SHOUHARDO project's operational area was not due to some broader, positive forces emanating from wider favorable economic or climatic trends in the country. This is because there were none. Project activities evidently not only shielded participant households from the adverse effects of the food price shock but also improved conditions to the point where stunting *declined*. Fewer and fewer children were becoming chronically malnourished over a time period when more children were becoming malnourished nationally. The total decline in stunting prevalence over the project period thus represents a valid lower-bound estimate of the magnitude of impact of its activities.

Four other pieces of information reinforce the finding that the decline in stunting prevalence in the SHOUHARDO project's operational area was not caused by external forces but by the project's interventions themselves. First, all four regions within the project area experienced reductions in stunting, as detailed in Figure 3. The largest reduction took place in Haor (18.3 percentage points) and the smallest in Mid Char (8.9 percentage points). Given that these regions are dispersed throughout the country and have widely varying environmental conditions, it is improbable that reductions counter-normal to the national trend would take place specifically in these areas only by chance.

Second, we find that households living close to, but not in, the project area display a declining trend in stunting, but far flatter than that of nearby SHOUHARDO project households. To look into this, the 2006 and 2010 HKI data sets were used to identify a sample of households located within project districts but not project sub-districts (upazilas). At the beginning of the project period these households had a stunting prevalence of 44.9; at its end it was 40.1.[27] This relatively small decline of 1.4 percentage points per year may be linked to the positive spillover effects from project interventions mentioned above.

Third, it is important to rule out the possibility that project households might have experienced greater declines in stunting than elsewhere simply because of the one trait they share in common: they are the poorest. Has stunting been declining more quickly among the poorest in Bangladesh than the general population? No. The annual decline in stunting prevalence among rural 6–24 month olds over 2000–2007 has been 1.3 percentage points per year. That for the poorest quintile of households has been only 0.3 percentage points.[28]

---

[26]   The cyclone did not directly affect households in the SHOUHARDO project area.

[27]   The sample sises for this analysis are 408 households from the 2006 survey (from four sub-districts of four districts) and 299 households from the 2010 survey (from 20 sub-districts of ten districts). Note that these samples are not a good comparison (or 'control') group for SHOUHARDO project households since the baseline value of stunting is far lower than that of SHOUHARDO households (44.9 versus 56.1 per cent). This difference is to be expected as project sub-districts were chosen based on their vulnerability status and the project households in them are the poorest of the poor, which is not the case for the comparison sample.

[28]   The 2000–2007 period was chosen for the comparison because comparable data on wealth quintiles are available in the 2000 and 2007 BDHSs. The quintiles, provided with the DHS data sets, are constructed from a wealth index generated using principal component analysis and data on household assets, including ownership of durable goods (such as televisions and bicycles) and dwelling characteristics (such as source of drinking water, sanitation facilities, and construction materials) (NIPORT *et al.* 2009).

HKI data collected between 1998 and 2005 exhibit the same pattern of lower reductions in malnutrition over time for poorer households (HKI 2006).

Figure 3 Reduction in prevalence of stunting among children 6–24 months in the SHOUHARDO project area, by region
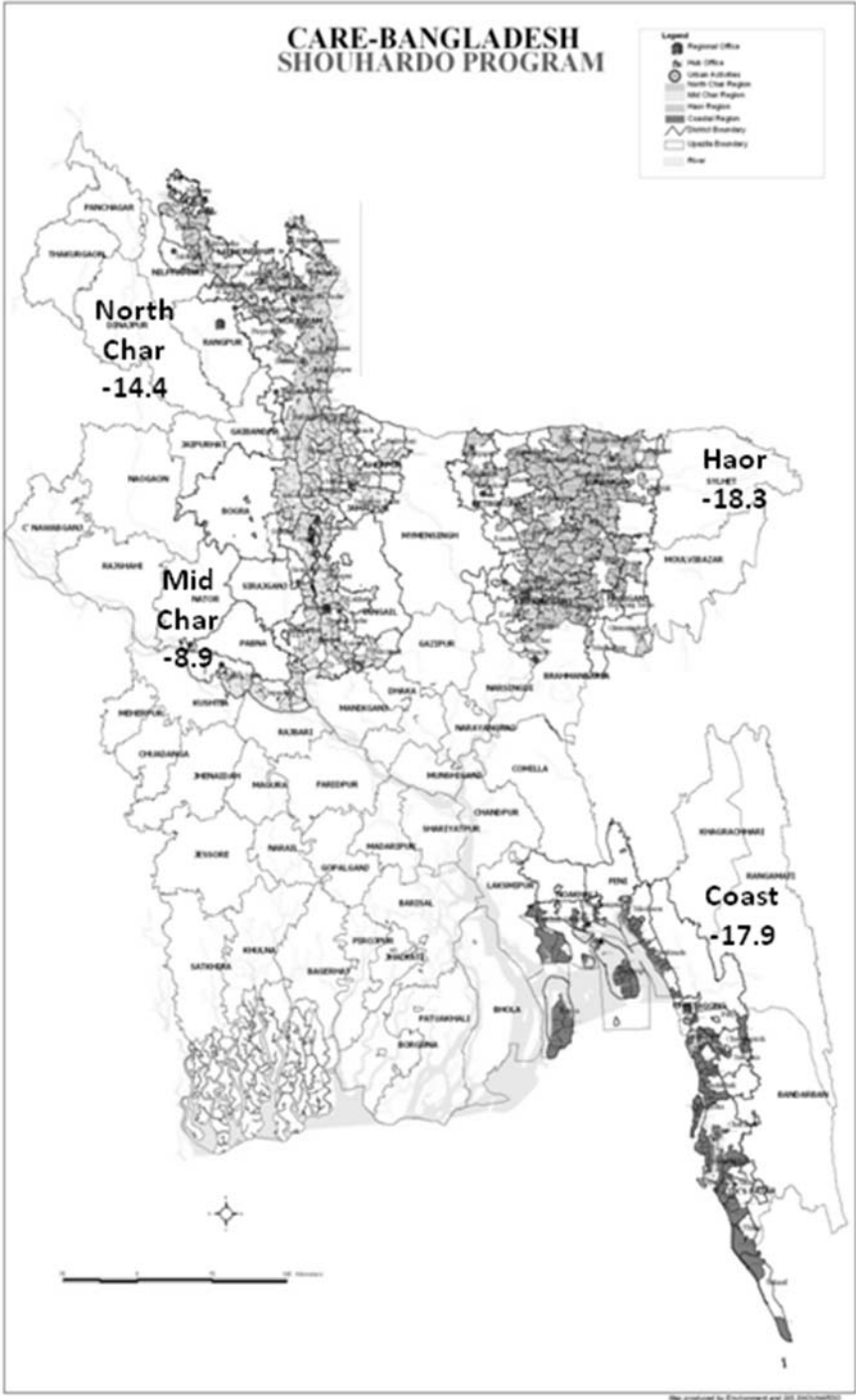
Fourth, the endline data confirm that very few households received assistance from the government or other NGOs during the project period. Indeed one of the criteria used for choosing the project areas was that no other major development activity was taking place by another organisation. Just over 10 per cent of households indicated that they had received

any aid from institutions or programs other than SHOUHARDO. The actual percent is lower since many of the institutions from which the aid was received were NGOs through which CARE implemented its programs. Further, since one of the goals of the project was to help households gain greater access to public services, in many of the cases household receipts of government assistance was ultimately due to SHOUHARDO project interventions.

4.2 Age trajectory of stunting in project area compared to nationally

Following the typical pattern for children from poor households in developing countries, in Bangladesh there is normally a steep increase in stunting as children age over the six month to 2 year-old range. This increase is associated with poor weaning practices and exposure to infectious disease. Continued high prevalences for older age groups are due to the initial growth failure at younger ages and possibly also poor household food access (Beaton *et al.* 1990).

DHS data collected in 2007 exhibit this typical pattern (see Figure 4). More specific to the age groups of interest to this evaluation, they show a rise in the stunting prevalence among rural Bangladeshi children from 29 per cent of all children 6–18 months old to 48 per cent of children 48–60 months old, an increase of nearly 20 percentage points (see Table 1). By contrast, there was virtually no increase in stunting prevalence among the children that had been exposed to SHOUHARDO project interventions. The prevalence rose by 1.2 percentage points across the two age groups, a difference that is not statistically significant (p=0.153). This small change is even more notable given that not all children in the 6–18 month group at baseline were exposed to the project's MCHN interventions for the full 18 month eligibility period (6–24 months) simply because they were not in the eligible age range for that long. For example, the 18 month olds were only exposed to project interventions for 6 months.

Figure 4 Age trajectory of stunting among 0–5 year olds in rural Bangladesh
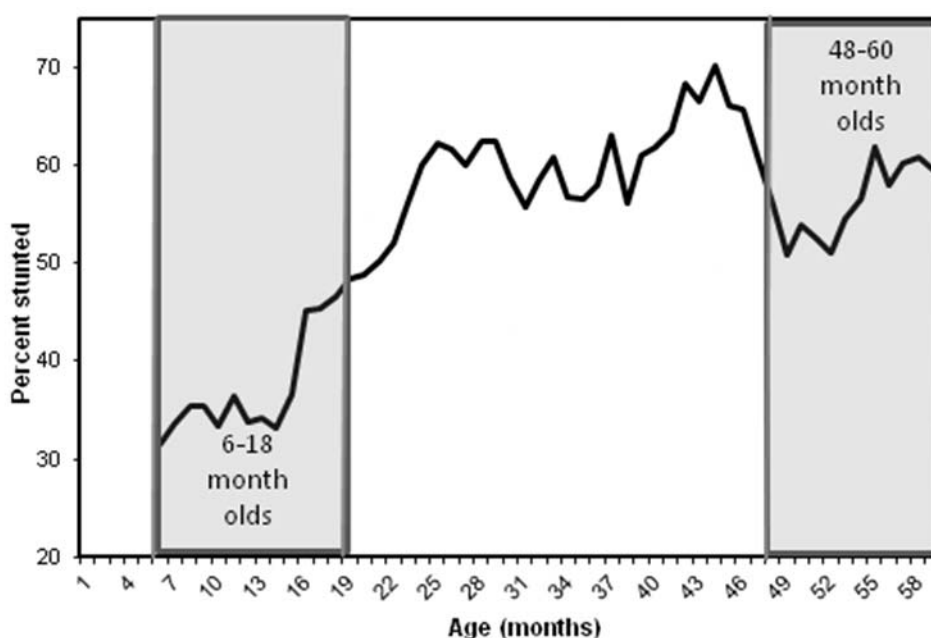
Source: Bangladesh Demographic and Health Survey 2007.

Table 1 Age trajectory of stunting among 0–5 year olds: comparison of SHOUHARDO project children with rural Bangladeshi children

	6–18 month olds	48–60 month olds	Increase (percentage points)
Rural Bangladeshi children	29.0	47.6	18.6
Project area children	51.4 (baseline)	52.6 (endline)	1.2

Sources: Rural Bangladeshi children: Bangladesh Demographic and Health Survey 2007; Project area children: baseline and August 2009 endline surveys.

We can deduce from this evidence that something happened to the children living in project households that prevented many of them from becoming stunted as they aged, another strong indication that the project's interventions plausibly led to a reduction in stunting.

4.3 Trends in the underlying determinants of child malnutrition

Table 2 reports baseline and endline values of indicators of food security, the quality of caring practices for children 6–24 months and their mothers, and in households' health environment quality.

Table 2 Changes in indicators of the underlying determinants of child malnutrition over the life of the project

Indicator	Baseline	Endline	Percent difference	p-value a/	
Food security					
Number of months sufficient food was accessed in the last year	5.5	8.9	61.8	0.000	***
Percent with 3 square meals a day most of the time or often in last year	33.5	74.1	121.2	0.000	***
Household dietary diversity score b/	5.0	6.3	25.7	0.000	***
Caring practices for children 6–24 months old					
Percent one year or older fully immunised	68.7	83.7	21.8	0.000	***
Percent receiving Vitamin A capsule in last 6 m	62.1	85.9	38.3	0.000	***
Percent 18–24 month olds breastfed	89.4	92.6	3.6	0.049	**
Percent given oral rehydration therapy during diarrhea	56.7	92.2	62.6	0.000	***
Percent of mothers washing hands before food preparation	60.3	94.3	56.4	0.000	***
Caring practices for mothers					
Percent having at least 3 antenatal visits during last pregnancy	15.8	57.8	265.8	0.000	***
Percent taking more food than usual during last pregnancy	6.2	53.9	769.4	0.000	***
Percent taking more daytime rest than usual during last pregnancy	25.3	46.1	82.2	0.000	***
Percent taking iron/folic acid during last pregnancy	27.1	80.5	197.0	0.000	***
Percent with Vitamin A supplementation within 1.5 months after delivery	9.4	78.9	739.4	0.000	***
Household health environment quality					
Percent with access to safe water for drinking, cooking, & washing	57.1	71.6	25.4	0.000	***
Percent with access to a sanitary latrine	13.8	54.6	295.7	0.000	***

a/ Stars indicate statistical significance at the 1% (***) or 5%(**) levels.

b/ Based on 24 hour recall for consumption of foods from 15 food groups. The endline estimate is from a project monitoring survey conducted in February 2009 (the same month as the baseline) as dietary diversity is dependent on seasonal factors.

Starting with food security, the data indicate that both the quantity and quality of households' diets improved considerably over the life of the SHOUHARDO project. The average number of months per year in which households report having had sufficient access to food rose from 5.5 at baseline to 8.9 by endline, nearly three and a half months. The percent of respondents reporting to have had 'three square meals' a day most of the time or often in the last year saw a substantial increase, from 34 to 74. The average household's dietary diversity score (an indicator of diet quality) increased by 26 per cent.

Turning to the quality of caring practices for children, the percent of children one year or older fully immunised increased from 69 to 84, ending higher than the national coverage rate (76 per cent) (NIPORT *et al.*, 2009). There was a considerable increase in the percent of children receiving Vitamin A supplementation, and the percent of children breastfed into their second year rose as well. With respect to diarrhea treatment and prevention, the percent of children given oral dehydration therapy during diarrhea increased dramatically, from 57 to 92. The percent of mothers who reported washing their hands before food preparation saw a hefty increase, rising from 61 to 94.

It is in the area of caring practices for mothers that the greatest improvements took place. The percent of mothers having at least three antenatal visits during pregnancy rose from 16 to 58, ending far higher than the national prevalence (32 per cent) (NIPORT *et al.* 2009). Other indications that there was pronounced improvement in the quality of care received by mothers during and following pregnancy are major increases in the percent of women taking more food and daytime rest than usual during pregnancy, receiving iron/folic acid supplementation during pregnancy, and receiving Vitamin A supplementation following delivery.

Finally, there have been significant improvements in the quality of households' health environments. The percent of households with access to safe water has increased from 57 to 72. Access to a sanitary latrine, which is crucial for preventing diarrhea, nearly quadrupled.

In sum, the data indicate that there have been considerable improvements in all three underlying determinants of child malnutrition over the SHOUHARDO project's implementation period. Given this evidence, it would be surprising if a significant reduction in malnutrition did *not* take place over that time as well.

5 Why did the project reduce stunting by so much?

There is one component of the SHOUHARDO project's targeting approach that we can assume – but not test – contributed to the decline in stunting achieved. Traditionally, MCHN interventions have been targeted to all children under five years old who are already malnourished. This is considered a 'recuperative' approach. Following recent findings from an MCHN-focused project in Haiti that 6 to 24 month olds are both most at risk of becoming malnourished and most responsive to nutritional interventions (Ruel *et al.* 2008; Donegan *et al.* 2010), the project's MCHN interventions, including food aid, were targeted only to children in this age range and their mothers. This 'preventive' approach was found in Haiti to lead to greater reductions in malnutrition than the recuperative approach.

However, as will be seen below, age targeting is probably not the only reason for the extraordinarily large reduction in stunting found among SHOUHARDO project households. Here, having established that the project did indeed have a large impact, we investigate the contributing roles of (1) household participation in the interventions other than MCHN; and

(2) targeting poor households rather than extending project interventions to all households in project villages.

5.1 Investigation of the added impacts of the sanitation, women's empowerment, and poverty alleviation interventions

Expected direction of selection bias
The PSM and MT-PSM methods employed for this investigation are based only on observable characteristics of project individuals, households, and districts. Selection bias due to unobservables is not accounted for. For this reason it is important to determine the likely direction of selection bias – whether due to targeting or self-selection – for each intervention in order to properly interpret the results. Table 3 lays out the information used to do so for each type of intervention. To render the analysis tractable, some of the interventions described in Section 1 are grouped into three broad intervention sets: MCHN (full participation), sanitation, and women's empowerment. This grouping is feasible due to their low coverage (see percent participation in Table 3), which leaves a large enough group of non-covered households from which to identify a control group. Since the poverty alleviation interventions as a group had wide coverage (88 per cent), it is necessary to analyse each individually.

With respect to targeting-based bias, information collected from project administrators reveals that in most cases there was either no targeting or geographical targeting was used to select the worst off areas for participation in interventions, which would lead to negative selection bias (and systematic underestimation of impact). In the case of the women's empowerment interventions, some were targeted within villages to women with 'leadership capacity' and 'community acceptance', which would lead to positive selection bias (systematic overestimation of impact).

With regard to self-selection of households into interventions, the expected benefits of participation likely motivated most to actually participate in those they were eligible for. However, personal endowments that ultimately influence children's nutritional status may have also affected their decision. Positive selection bias may have been at work if more motivated, skilled or socially connected persons were more likely to participate. Negative selection bias may have been at work if needier households were more likely to participate. The overall expected directions of selection bias are given in the final column of Table 3.

Table 3 Expected direction of selection bias, by intervention

Intervention	Percent of households	Bias due to targeting	Bias due to self-selection	Overall direction of bias
Mother and Child Health and Nutrition (MCHN), Full participation	33.0	None. Universal coverage extended to project participants.	While more motivated/capable mothers may be more likely to fully participate, the fact that those from worse-off households are in greater need of assistance in the form of a direct food transfer means that negative bias dominates.	Negative.
Sanitation	18.6	Negative. Targeted to households without sanitation facilities.	None. Households have no reason to refuse free assistance.	Negative.
Women's empowerment	24.8	Negative &positive. Geographical targeting to areas where women are most disadvantaged. a/ Within villages, VDCs chose members of EKATA groups and school PTAs with preference to those with leadership capacity and community acceptance.c/	Positive due to motivation/capability/social connectedness (although most women chosen for participation by the VDCs did in fact participate, they were consulted about their participation as part of the decision making process).	Indeterminant (Although not likely to be strong in either direction).
Poverty and food insecurity alleviation				
Field crop production/fisheries	36.4	None. b/	While households with more motivated/capable individuals would be more likely to participate, so too would the most needy.	Indeterminant.
Homestead gardening and livestock rearing	46.3	None. b/	As above.	Indeterminant.
Income generating activities	36.7	None. b/	As above.	Indeterminant.
Food/cash-for-work	10.5	Negative. Geographical targeting to areas needing flood and sanitation infrastructure. Targeted to the most needy households.	Negative. More likely to be taken up by the most needy.	Negative.
Savings groups	31.6	Negative. Targeted to areas with low presence of microfinance institutions, low access of women to income generating activities, and presence of an EKATA women's empowerment group.	Positive. More motivated/capable/socially connected individuals and those with more resources to put aside for future uses more likely to have participated.	Indeterminant.

a/ Factors taken into account were 1) women's mobility and access to information & education; 2) women's access to income generating activities; 3) rates of violence against women; 4) social conservativeness; 5) low engagement of other service providers to combat violence against women.
b/ Households were initially classified into four 'core occupational groups' based on their main occupations, apart from wage labor, at the start of the project: agricultural production, fisheries, homestead production, and self-employed income generating activities. They were eligible to participate in the first three of the poverty alleviation interventions based on these classifications, but no targeting within the classifications was undertaken. In the PSM analyses the initial classifications are partially controlled for with the inclusion of households' main occupations as a characteristic on which matching is based.
c/ In the case of early childhood development the only preference given was to girl children.

Propensity score estimations

Three groups of independent variables representing characteristics that are likely to affect both the probability of participating in an intervention as well as children's nutritional status – but unlikely to have been affected themselves by an intervention – are controlled for in the probit propensity score estimations. The first group contains dummy variables for participation in the other intervention sets. For example, when estimating the propensity scores for participation in sanitation, participation in the MCHN, women's empowerment and poverty alleviation interventions is also controlled for. This is necessary for isolating the independent (stand alone) impact of each intervention set. The second group is key determinants of nutritional status that are not intended to be influenced by the SHOUHARDO project: age and sex of the child; age and educational achievement of the child's mother; age, gender and main occupation of the household head; household size and age-sex composition; an indicator variable for whether or not the household was classified as 'extreme poor' (as opposed to 'poor') before project activities started; each household's region of residence and, in addition, two village-level variables that possibly influence participation, the number of households in the village and the percent of households participating in the SHOUHARDO project (that is, that are classified extreme poor or poor). Finally, to minimise targeting-based selection bias, for some interventions we include a group of characteristics that may have potentially influenced the allocation of project interventions to various project districts. These characteristics, measured at the district level using baseline (not endline) survey data, are:

- Sanitation: Percent of households with access to safe water and sanitary latrines;
- Women's empowerment: Score for women's involvement in major decisions,[29] percent of women earning cash income, percent of school-aged children attending school, and percent of adults literate;
- Food/cash-for-work: Percent of households experiencing a disaster in the last year and in the last five years, and percent of households with access to safe water and sanitary latrines;
- Savings groups: The characteristics listed under 'women's empowerment' above and an indicator of the availability of microfinance institutions, the percent of households in the district that participate in a credit group.

Results of the probit regressions for the MCHN (full participation), sanitation, women's empowerment, and an illustrative poverty reduction intervention, agricultural production/fisheries, are given in Appendix Table 1.

Common support

The common support condition is strongly satisfied, as illustrated in Appendix Figure 1, which shows the propensity score distribution of participating versus control group households for some of the interventions. For all interventions there are ample non-participating households with propensity scores close by in the distribution with which to be matched, with the exception of a few having very high propensity scores. Across all ATT estimates the mean percent of participating households excluded from the analysis due to lack of common support is 1.5 (maximum: 3.5).

PSM results for the independent impacts of interventions

The PSM results for the independent (stand alone) impacts of the interventions sets on children's height-for-age z-scores are presented in the upper panel of Table 4. Consistent with the chi-squared p-value for the test of matching quality (all =1.0, see column E), after matching there are only two cases in which there is any significant difference in the mean values of the characteristics on which matching is based between the participant and

---

[29]　Calculated based on the ability of women to take part in decisions regarding the buying or selling of major household assets and jewelery, the use of loans or savings, and expenses for children's education.

matched control groups.[30] The sizes of the participant and control groups are given in columns C and D.

The estimated ATT for full (versus partial) participation in MCHN is 0.114 z-scores. Given that this estimate is likely affected by negative selection bias, its sign, magnitude and statistical significance confirm that greater participation in the direct nutrition interventions leads to better long-term nutritional status. The raw results indicate very little if any impact of the sanitation and poverty alleviation interventions on nutritional status when they are not combined with MCHN interventions. Because of the negative selection bias plaguing the ATT estimates for the sanitation interventions and indeterminant direction for the poverty alleviation interventions, no strong conclusion can be reached regarding this finding.

The PSM results indicate that the women's empowerment interventions, by contrast, likely had quite a strong impact even in the absence of the MCHN interventions. The estimated ATT is 0.226, which amounts to a reduction in stunting prevalence for participants of just over six percentage points.[31] Note that any selection bias affecting this estimate is expected to be small (see Table 3). Note also that spillover of the positive nutritional benefits of this intervention to other households located in the vicinity, which is a real possibility, would lead to underestimation of the ATT.

Multiple-treatment PSM tests for impact synergies

The multiple treatment PSM tests of synergies from combining the interventions addressing structural causes with MCHN interventions are presented in the bottom panel of Table 4. Only the ATT results for interventions demonstrating that such synergies may indeed exist are presented.[32] To help interpret the results, the implied equivalent reductions in stunting are given in Figure 5. As for the estimates of the independent impacts of the interventions, matching quality is strong; after matching, there are few remaining differences in the characteristics on which matching is based across the participant and control groups.[33]

---

[30] For field crop production/fisheries promotion and income generating activities there remain significant differences across the two groups for the percent of households participating in homestead food production promotion, a difference of approximately five percentage points.

[31] The equivalent impact on stunting is calculated by computing, for both participant and control groups, the area to the left of -2 z-scores under the standardised-normal HAZ distribution function with (unstandardised) mean being the estimated HAZ from the PSM analysis. The difference in the two stunting proportions multiplied by 100 is the estimated impact on stunting.

[32] In the case of food/cash-for-work, synergies were detected but the sample size for the group of households participating in both MCHN (fully) and food/cash-for-work was too small for making inferences (N=95).

[33] Remaining differences that are statistically significant are: 1) For 'women's empowerment only': participation in savings groups (participants 25 per cent; non-participants 20 per cent) and the percent of households residing in the 'Coast' region (25.8 versus 20.8); and 2) for field crop production/fisheries promotion: participation in sanitation (21.8 versus 16.2 per cent) and the percent of males in the 15-64 year age group (22.8 versus 23.9 per cent).

Table 4 Propensity score matching estimates of average treatment effects on the treated of project interventions on height-for-age z-scores of 6–24 month old children

	Average treatment effect on the treated (ATT)	z-stat		Number of observations		Chi-squared p-value for test of matching quality
				Partic-ipants	Controls	
	(A)	(B)		(C)	(D)	(E)
Estimates of the independent impacts of intervention sets						
(Kernel matching)						
MCHN (Full participation)	0.128	2.07	**	915	1,636	1.000
Sanitation	0.114	1.27		489	2,056	1.000
Women's empowerment	0.226	3.09	***	638	1,912	1.000
Increasing food production and incomes						
Crop production/fisheries	0.081	1.28		893	1,644	1.000
Homestead food production	0.003	0.05		1,154	1,363	1.000
Income generating activities	-0.026	-0.44		887	1,659	1.000
Food/cash-for-work	0.087	0.87		255	2,298	1.000
Savings groups	0.085	1.03		792	1,754	1.000
Tests for synergies between MCHN and other intervention sets						
(Multiple treatment kernel matching)						
MCHN and sanitation						
MCHN only	0.134	1.97	**	797	1,240	0.987
Sanitation only	0.127	1.13		275	1,240	0.989
Both	0.248	2.15	**	211	1,240	1.000
MCHN and women's empowerment						
MCHN only	0.096	1.27		759	1,146	0.223
Empowerment only	0.242	2.19	**	371	1,146	0.759
Both	0.313	2.72	***	255	1,146	0.947
MCHN and crop production/fisheries promotion						
MCHN only	0.052	0.68		633	1,007	0.842
Crop production/fisheries only	0.052	0.62		510	1,007	0.570
Both	0.298	2.50	**	397	1,007	0.985

Notes: Stars represent statistical significance at the 1%(***), 5%(**) and 10%(*) levels. For kernel matching, an epanechnikov kernel function is used. Standard errors (and thus the reported z-statistics) are estimated by bootstrapping with 100 repetitions.

Source: Authors' calculations.

Figure 5 Multiple treatment propensity score matching estimates of synergistic impacts from combining interventions on stunting among 6–24 month olds
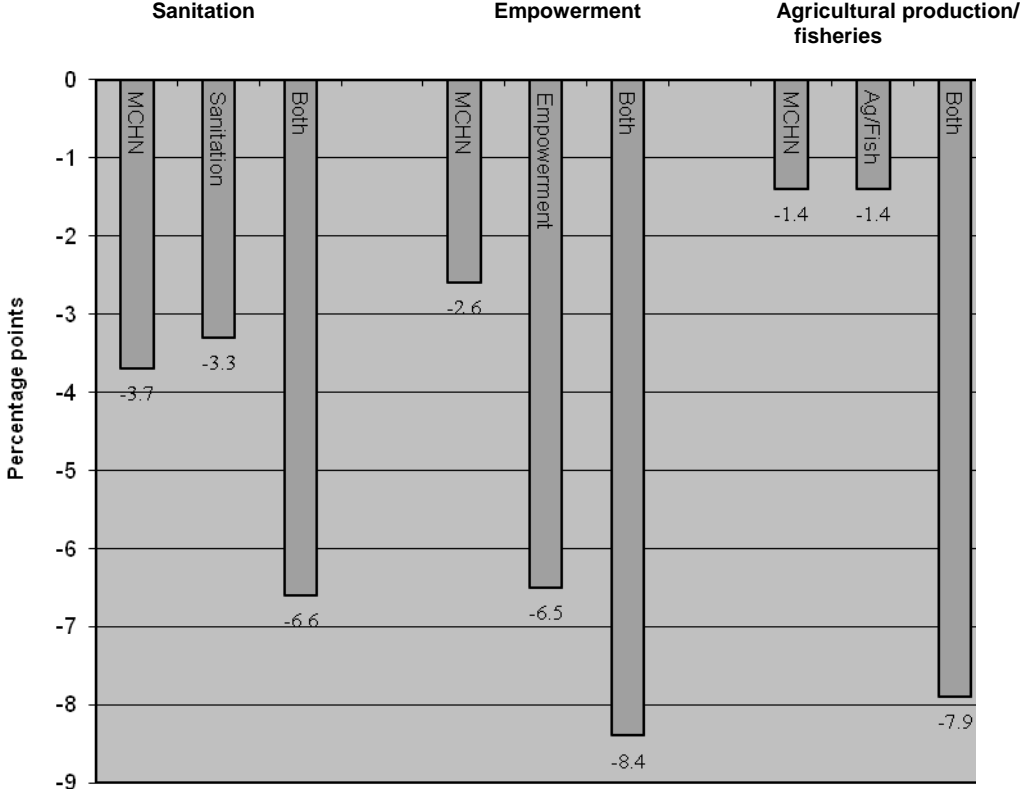
Source: Authors' calculations.

With regard to sanitation, compared to the households that participated in sanitation interventions but only partially in MCHN interventions, the estimated impact on the stunting prevalence of participating in *both* is double (Figure 5). Given that any remaining selection bias in these estimates is likely to be negative, they suggest strong synergies between the project's direct nutrition and sanitation interventions.

The data also give evidence of synergies between the MCHN and women's empowerment interventions. The results suggest that the reduction in stunting for the sub-set of project households participating in both interventions was far greater than for those that participated in MCHN (full participation) alone and even greater than those that participated in the empowerment interventions alone. The validity of this result is strengthened by the fact that selection bias is not likely having a large influence on the estimates.

Turning finally to the interventions aimed at reducing poverty and food insecurity, only that aimed at increasing food production through field crop production and fisheries showed evidence of impact synergies with the MCHN interventions. The implied very large reduction in stunting when the interventions are combined compared to the small reductions when they are not implies strong synergies that are unlikely driven by any positive selection bias alone.

5.2 The role of pro-poor targeting

It seems logical that children are more likely to benefit from health and nutrition interventions the poorer is the household they live in simply because they are more likely to be suffering from the conditions leading to malnutrition, such as poor diets and infectious disease. A

recent review of the impacts of interventions aimed at reducing child malnutrition reported that most evaluations found children from poorer households to indeed benefit more. However, some interventions were found to benefit richer households more, and for some there was no difference (World Bank 2010b). The degree to which the poor benefit more may depend on the types of interventions, including whether they are direct nutrition interventions and/or address structural causes, the context in which they are implemented, and the extent of socio-economic differentiation.

As mentioned above, only the poorest households within the geographical area of the SHOUHARDO project were selected to participate in its activities. Would the project have had lower overall impact if the richer households had also been included, i.e., if the project employed universal coverage? Having already demonstrated that the large reduction in stunting is in fact likely to be due to project interventions, here we compare the change in stunting prevalence over the life of the project for the two economic groups included in the project, the poor and extreme poor, to gain some insight into this question.[34] Note that intervention-by-intervention there is no difference in the participation rates of the extreme poor and poor groups of households except in the case of sanitation, for which 23 per cent of extreme poor households participated and only 16.6 per cent of poor households.

Figure 6 shows that the reduction in the prevalence of stunting over the life of the project was far greater for the extreme poor project households than the poor (21.3 versus 12.7 percentage points). These results confirm that the overall reduction would have been lower had the richer households been included as project participants and that the SHOUHARDO project's strategy of targeting the very poorest, rather than practicing universal coverage, was one of the reasons that it was able to reduce child malnutrition by so much.

Figure 6 Change in the prevalence of stunting among 6–24 month olds between the baseline and endline surveys, by poverty status
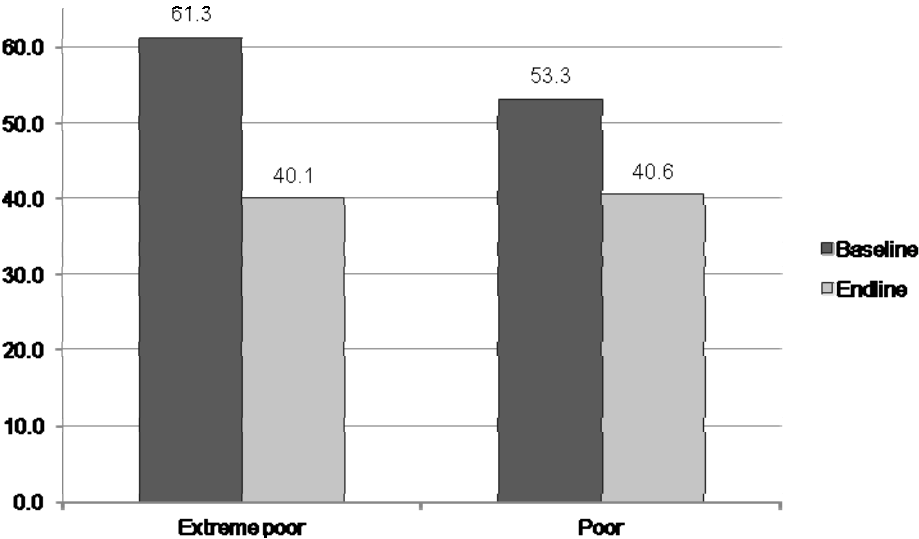
Notes: Prevalences are calculated using the data collected in the project baseline (February 2006) and endline (November 2009) surveys.

---

[34]   Because of insufficient sample size it was not possible to investigate the differential impacts of individual project interventions on the poor and extreme poor using PSM.

6 Conclusion

This paper has presented a body of plausible evidence that the SHOUHARDO project – the first large-scale nutrition-oriented project using the rights-based, livelihoods approach – had a large impact on child malnutrition. At the same time it has given some insight into why the project had such an impact, which is increasingly considered crucial information for rendering evaluations policy relevant and judging their external validity (Ravallion 2008; White 2009; IIIE 2011a).

To summarise:

(1) The stunting prevalence among project participants fell by an unusually large 16 percentage points over a three-and-a-half year period during which stunting was stagnant in Bangladesh as a whole and even increasing for some time due to a major food price crisis and adverse weather conditions;

(2) The normal substantial increase in the stunting prevalence as young children age over the 0–5 year range did not occur at all for the group of children living in project households, indicating that project interventions prevented many children from becoming malnourished;

(3) Substantial improvements in all of the underlying determinants of child malnutrition, including food security, the quality of caring practices for children and women, and health environment quality, took place over the life of the project, improvements that would in turn be expected to lead to substantial improvements in children's long term nutritional status;

(4) The project's women's empowerment interventions were found to have a strong independent impact on stunting, and the sanitation, women's empowerment, and one poverty alleviation intervention were found to have synergistic impacts with direct nutrition interventions. These findings confirm that the project activities addressing structural causes contributed to the reduction in malnutrition and, more generally, that the use of the rights-based, livelihoods programming approach was instrumental in bringing it about;

(5) The reduction in stunting was far greater for extreme poor than poor project households, evidence that the use of pro-poor targeting rather than universal coverage also facilitated the reduction in stunting.

While each of these pieces of information alone would not likely be entirely convincing, together they provide solid evidence of the project's impact.

A valuable lesson for future development projects aimed at reducing child malnutrition emerges from the SHOUHARDO experience. That is, combining direct nutrition interventions, such as the 13 proposed by the Scaling Up Nutrition initiative, with those that address structural causes – at the same time and for the same households – has the potential to accelerate reductions in child malnutrition at a rate far greater than can be expected from direct nutrition interventions alone. The findings also add to the evidence that in many settings carefully targeting the poor is a more efficient way of reducing malnutrition that extending universal coverage.

In line with the call for more diverse and holistic impact evaluation methodologies, we employed a mixed methods quantitative approach based on simple descriptive and more complex statistical methods, both non-experimental and quasi-experimental, triangulating the

information to arrive at the overall conclusion. Key to interpreting the quantitative results were (1) close knowledge of what was happening country wide during the project period; and (2) information on potential factors affecting selection of project participants into particular project interventions. Some limitations were encountered, for example we could not determine whether some interventions contributed to the decline in stunting because of insufficient information on the magnitudes of the sources of positive and negative selection bias. Such limitations and, more generally, lack of an established protocol with standardised procedures and clear-cut guidelines for what constitutes 'impact' may lead to some discomfort. Yet dealing with ambiguity is the reality of most evaluations of development activities, which require strict objectivity combined with flexibility and creativity on the part of the evaluator.

The evidence presented was arrived at in the absence of randomisation, a non-project control group, or panel data, some combination of which is required to achieve precise estimation of the magnitude of impact of a project. Despite the robustness of the evidence presented, this fact means that the results could be – and have been – dismissed by some. What are the implications for forwarding the cause of development effectiveness? If such draconian requirements are used to judge the validity of impact evaluations then the important lessons learned from projects like SHOUHARDO will not be disseminated for consideration in other settings. Taken to the extreme, the SHOUHARDO project would risk discontinuation. As a result a large proportion of poor households living in adverse conditions in Bangladesh would not benefit from the successful set of interventions being promoted by the project. Fortunately, this did not happen, and SHOUHARDO II, following roughly the same model as SHOUHARDO I, is now in its start-up phase.

Writes Lin (2011): '…the ultimate test of relevance is not the desirability of a certain research methodology, but the importance of the question that the research asks and the policy insights that such research generates'. In conclusion, given the valuable policy lessons generated and the rigor with which this illustrative evaluation was conducted, the experience of the SHOUHARDO project merits solid standing in the knowledge bank of development effectiveness.

Appendices

Appendix Table 1 Probit propensity score model estimation for participation in selected project interventions

Variable	MCHN (full participation)		Sanitation		Women's empowerment		Crop production /fisheries	
	Coeff-icient	z-stat	Coeff-icient	z-stat	Coeff-icient	z-stat	Coeff-icient	z-stat
Participation in other interventions								
MCHN (Full participation)	--	--	0.223	**3.48**	0.045	0.69	0.234	**4.09**
Sanitation	0.225	**3.31**	--	--	-0.152	**-1.94**	0.366	**5.28**
Women's empowerment	0.068	1.03	-0.169	**-2.22**	--	--	0.131	**1.98**
Crop production/fisheries	0.230	**3.97**	0.365	**5.58**	0.152	**2.36**	--	--
Homestead food production	0.058	1.06	0.280	**4.48**	0.004	0.07	-0.341	**-6.15**
Income generating activities	0.039	0.68	0.277	**4.22**	0.121	**1.89**	-0.613	**-10.40**
Food/cash-for-work	0.066	0.75	0.220	**2.25**	0.230	**2.41**	0.168	**1.89**
Savings groups	-0.171	**-2.81**	0.286	**4.23**	0.513	**7.86**	0.400	**6.64**
Child and household characteristics								
Age of child	0.013	0.34	0.022	0.52	0.015	0.36	0.020	0.54
Age of child-squared	0.001	0.47	-0.001	-0.40	0.000	0.03	-0.001	-0.75
Whether child is a girl	-0.052	-0.73	0.041	0.50	-0.156	**-2.08**	0.056	0.78
Age of mother	0.005	0.71	-0.001	-0.17	0.007	0.89	0.003	0.44
Mother's education: none a/								
Mother's education: primary	0.092	1.58	0.025	0.37	0.135	**2.05**	0.110	**1.85**
Mother's education: secondary	0.130	1.52	0.024	0.24	0.188	**1.89**	0.045	0.51
Whether household is headed by a female	0.080	0.42	-0.101	-0.48	-0.176	-0.75	0.153	0.76
Age of household head	-0.002	-0.49	0.001	0.23	-0.001	-0.28	0.006	**1.72**
Primary occupation: Farming a/								
Agricultural laborer	0.165	**2.29**	-0.006	-0.07	-0.054	-0.67	-0.433	**-5.99**
Non-agricultural laborer	0.082	0.89	-0.225	**-2.09**	-0.136	-1.33	-0.536	**-5.80**
Salaried employment	0.283	**2.16**	0.038	0.25	0.010	0.07	-0.336	**-2.46**
Trading and self-employment	0.295	**3.49**	0.054	0.57	-0.014	-0.14	-0.471	**-5.52**
Unpaid household work	0.719	**2.40**	0.249	0.80	0.113	0.32	-0.626	**-2.00**
Other	0.260	**2.12**	0.012	0.08	0.143	1.05	-0.526	**-4.13**

Household size	-0.040	**-1.93**	0.083	**3.79**	0.116	**5.18**	-0.046	**-2.18**
Percent of females 0-15 years a/								
Percent of females 15-64	0.006	1.56	0.010	**2.38**	-0.023	**-5.63**	-0.008	**-2.02**
Percent of females 64 or more	-0.005	-0.64	0.014	1.58	-0.019	**-2.17**	-0.006	-0.79
Percent of males 0-15 years	0.000	-0.09	0.004	1.43	-0.004	-1.62	0.001	0.53
Percent of males 15-64	0.002	0.56	0.006	1.34	-0.033	**-7.33**	0.003	0.66
Percent of males 64 or more	0.017	**2.09**	-0.006	-0.67	-0.010	-1.15	0.010	1.19
Economic status (pre-intervention)								
Poor a/								
Extreme poor	0.012	0.2	0.311	**4.67**	-0.166	**-2.45**	0.017	0.28
Number hhs in village of residence	0.000	1.49	-0.001	**-2.31**	0.000	-1.17	0.000	1.04
Percent of hhs in village participating the SHOUHARDO project	0.000	-0.28	0.001	0.92	-0.002	**-2.18**	-0.001	**-2.05**
Region of residence North Char a/								
Mid Char	-0.088	-1.14	-0.260	**-2.13**	-0.005	-0.03	0.498	**6.02**
Haor	-0.555	**-6.69**	-0.175	-1.19	-0.328	**-2.23**	0.501	**5.82**
Coast	-0.219	**-2.77**	0.537	**4.12**	-0.020	-0.18	0.566	**6.78**
District-level targeting variables (measured pre-intervention)								
Percent with access to safe water	--	--	0.004	**1.74**	--	--	--	--
Percent with access to a sanitary latrine	--	--	-1.411	**-3.48**	--	--	--	--
Score for women's involvement in major decisions	--	--	--	--	-0.336	-0.70	--	--
Percent of women earning cash income	--	--	--	--	-0.007	-0.56	--	--
Percent of school-aged children attending school	--	--	--	--	-0.034	**-3.36**	--	--
Percent of adults literate	--	--	--	--	0.023	**2.10**	--	--
Number of observations	2,553		2,553		2,553		2,553	
Psuedo R-Squared	0.053		0.109		0.154		0.107	

Notes: Bolded z-statistics signify that the coefficient is significant at least at the 10% level.
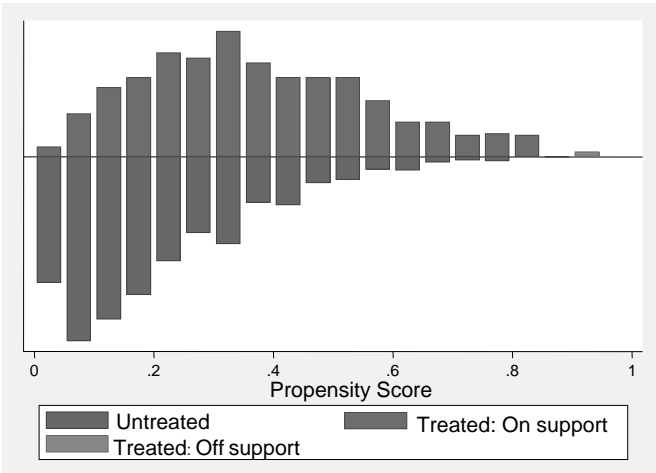a/ Reference category

Appendix Figure 1 Common support: propensity scores of treated and control households for selected interventions
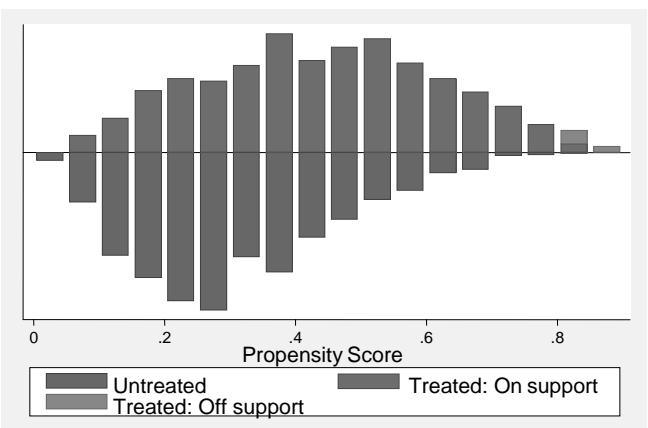
Sanitation
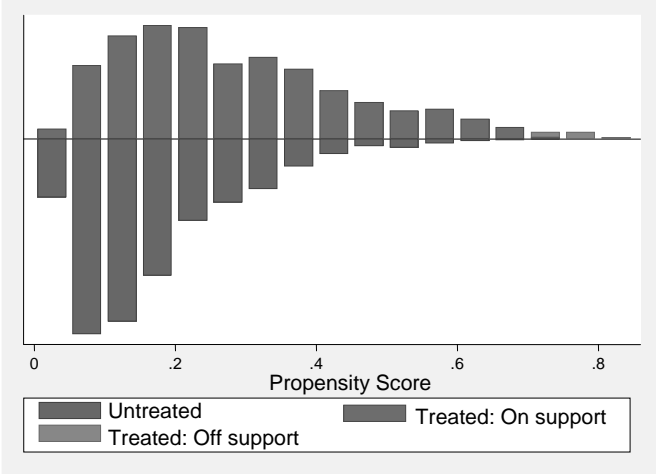
Women's empowerment

Agricultural production/Fisheries

References

American Heritage Dictionary (2009) *Dictionary of the English Language,* fourth education, Houghton Mifflin Company

Banerjee, Abhijit (2007) *Making Aid Work,* Cambridge, Mass: MIT Press

Barrett, Christopher B. and Carter, Michael R. (2010) 'The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections', *Applied Economic Perspectives and Policy* 32.4: 515–48

Beaton, G.; Kelly, A.; Kevany, J.; Martorell, R. and Mason, J. (1990) *Appropriate Uses of Anthropometric Indices in Children,* ACC/SCN State-of-the-Art Series, Nutrition Policy Discussion Paper 7, ACC/SCN: Geneva, December

Becker, Sascha O. and Egger, Peter H. (2007) *Endogenous Product versus Process Innovation and a Firm's Propensity to Export,* CESIFO Working Paper 1906, IFo Munich, Germany: Institute for Economic Research, University of Munich

Blundell, Richard; Dearden, Lorraine and Sianesi, Barbara (2004) *Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS,* London: Center for the Economics of Education, London School of Economics

Caliendo, Marco and Kopeinig, Sabine (2008) 'Some Practical Guidance for the Implementation of Propensity Score Matching', *Journal of Economic Surveys* 22.1: 31–72

Chambers, Robert (2009) 'So that the Poor Count More: Using Participatory Methods for Impact Evaluation', *Journal of Development Effectiveness* 1.3: 243–6

CGD (Center for Global Development) (2006) *When Will We Ever Learn? Improving Lives through Impact Evaluation,* Washington, D.C.: Center for Global Development

Cogill, Bruce (2003) *Anthropometric Indicators Measurement Guide,* Washington, D.C.: Food and Nutrition Technical Assistance Project, Academy for Educational Development

Corning, Peter A. (2007) 'Synergy Goes to War: An Bioeconomic Theory of Collective Violence', *Journal of Bioeconomics* 9.2

Deaton, Angus (2010) 'Instruments, Randomization, and Learning about Development', *Journal of Economic Literature* 48.2: 424–55

—— (2007) *Forum: Angus Deaton Making aid work* Cambridge, Mass: MIT Press

de Onis, Mercedes; Cutberto Garza, Cesar G.; Victora, Maharaj K. Bhan and Kaare R. Norum (guest eds) (2004) 'The WHO Multicentre Growth Reference Study (MGRS): Rationale, Planning, and Implementation', *Food and Nutrition Bulletin* 25 (supplement 1): S3–S84

Donegan, Shannon; Maluccio, John A.; Myers, Caitlin K.; Menon, Purnima; Ruel, Marie T. and Habicht, Jean-Pierre (2010) 'Two Food-assisted Maternal and Child Health Nutrition Programs Helped Mitigate the Impact of Economic Hardship on Child Stunting in Haiti', *Journal of Nutrition* 140: 1139–45

Drinkwater, Michael (2001) 'The Challenge of Linking Livelihood and Rights Approaches to Human Development', mimeo, Atlanta: CARE International

Duflo and Kremer (2003) 'Use of Randomization in the Evaluation of Development Effectiveness'*,* paper prepared for the World Bank Operations Evaluation Department (OED) Conference on Evaluation and Development Effectiveness, Washington, D.C.

EES (European Evaluation Society) (2007) *The Importance of a Methodologically Diverse Approach to Impact Evaluation – Specifically with Respect to Development Aid and Development Interventions,* The Netherlands: ESS

Elbers, Chris; Gunning, Jan Willem and Kobus de Hoop (2009) 'Assessing Sector-wide Programs with Statistical Impact Evaluation: A Methodological Proposal', *World Development* 37.2: 513–20

Esrey, Steven (1996) 'Water, Waste, and Well-being: A Multicountry Study', *American Journal of Epidemiology* 143.6: 608–23

Frankenberger, Timothy R. (2001) 'The Role of Non-Government Organizations in Promoting Nutrition and Public Health', paper presented at the 17th International Congress of Nutrition, Vienna, Austria, 26–31 August 2001

Frankenberger, Timothy R.; Drinkwater, Michael and Maxwell, Daniel (2000) *Operationalizing Household Livelihood Security: A Holistic Approach for Addressing Poverty and Vulnerability,* Atlanta: CARE, International

Gillespie, Stuart; Mason, John and Martorell, Reynaldo (1996) *How Nutrition Improves. Nutrition Policy,* Discussion Paper 15, ACC/SCN State-of-the-Art Series, United Nations-Administrative Committee on Coordination – Subcomittee on Nutrition

Handa, Sudhanshu and Maluccio, John A. (2010) 'Matching the Gold Standard: Comparing Experimental and Nonexperimental Evaluation Techniques for a Geographically Targeted Program', *Economic Development and Cultural Change* 58.3: 415–47

HKI (Helen Keller International) (2011) Results of primary data analysis provided by Jillian Waid, May 2011

—— (2010) 'The Food Security and Nutrition Surveillance Project Round 1: January 2010–April 2010 Preliminary Results', *Nutritional Surveillance Project Bulletin* 7, Dhaka, Bangladesh: Hellen Keller International

—— (2006) 'Trends in Child Malnutrition, 1990 to 2005: Declining Rates at National Level Mask Inter-regional and Socioeconomic Differences', *Nutritional Surveillance Project Bulletin* 19, Dhaka, Bangladesh: Hellen Keller International

Ianotti, Lora and Gillespie, Stuart (2002) *Successful Community Nutrition Programming: Lessons from Kenya, Tanzania, and Uganda,* LINKAGES: Breastfeeding, LAM and Related Complementary Feeding and Maternal Nutrition Program, Washington, D.C., the Regional Centre for Quality of Health Care, Makerere University, Uganda, and UNICEF, New York

IIIE (International Initiative for Impact Evaluation) (2011a) *Principles for Impact Evaluation,* http://www.3ieimpact.org/strategy/pdfs/principles%20for%20impact%20evaluation.pdf (accessed May 2011)

—— (2011b) *Impact Evaluation (IE) Studies Database: Quality Standards,* http://www.3ieimpact.org/database_of_impact_evaluations.html (accessed June 2011)

Imbens, G. (2001) 'The Role of the Propensity Score in Estimating Dose-response Functions', *Biometrika* 87.3: 706–10

IRIN (2010) *Bangladesh: Towards 'Sanitation for All by 2010',* IRIN: Humanitarian news and analysism, a project of the UN Office for the Coordination of Humanitarian Affairs, www.irinnews.org/Report.aspx?ReportID=77094

J-PAL (Abdul Latif Jameel Poverty Action Lab) (2011a) *What is Randomization?,* www.povertyactionlab.org/methodology/what-randomization (accessed June 2011)

—— (2011b) *Overview,* www.povertyactionlab.org/methodology (accessed June 2011)

—— (2011c) *Why Randomize?,* www.povertyactionlab.org/methodology/why-randomize (accessed June 2011)

Jennings, Joan and McCaston, M. Catherine (2007) 'Evaluating for Synergy: Improving Measurement of the Impact of Combined Efforts', mimeo, Atlanta: CARE International

Jones, Harry (2009) 'The "Gold Standard" is Not a Silver Bullet for Evaluation', *Opinion* London: Overseas Development Institute

Jones, Nicola; Jones, Harry; Steer, Liesbet and Datta, Ajoy (2009) *Improving Impact Evaluation Production and Use,* Overseas Development Institute Working Paper 300, London: Overseas Development Institute

Khagram, Sanjeev; Thomas, Craig; Lucero, Catrina and Mathes, Subarna (2009) 'Evidence for Development Effectiveness', *Journal of Development Effectiveness* 1.3: 247–70

Khandker, Shahidur R.; Koolwal, Gayatri B. and Samad, Hussain A. (2010) *Handbook on Impact Evaluation: Quantitative Methods and Practices,* Washington, D.C.: The World Bank

Lawrey, Steven (2010) *When too Much Rigor leads to Rigor Mortis: Valuing Experience, Judgment and Intuition in Nonprofit Management,* blog, Hausercenter.org

Lechner, M. (2001) 'Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption', in M. Lechner and F. Pfeiffer (eds), *Econometric Evaluation of Labour Market Policies,* Heidelberg: Physica/Springer

Leeuw, Frans and Vaessen, Jos (2009) *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation,* The Network of Networks on Impact Evaluation

Leuven, E. and Sianesi, B. (2003) *PSMATCH2: Stata Module to Perform full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing,* Version 3.1.5 2 May 2009

Levinson, F. James; Lorge Rogers, Beatrice; Hicks, Dristin M.; Schaetzel, Thomas; Troy, Lisa and Young, Collette (1999) *Monitoring and Evaluation: A Guidebook for Nutrition Project Managers in Developing Countries,* Washington, D.C.: World Bank

Lin, Justin Yifu (2011) *Watermelons versus Sesame Seeds,* submitted Thursday 6 June 2011, http://blogs.worldbank.org/developmenttalk/watermelons-vs-sesame-seeds

Moreno-Serra, Rodrigo (2009) *Health Programme Evaluation by Propensity Score Matching: Accounting for Treatment Intensity and Health Externalities with an Application to Brazil,* HEDG Working Paper 09/05 Health, econometrics and data group, York, United Kingdom: the University of York

Moser, Caroline and Norton, Andy (2001) *To Claim our Rights: Livelihood Security, Human Rights and Sustainable Development,* London: Overseas Development Institute

National Institute of Population Research and Training (NIPORT), Mitra and Associates, and Macro International (2009) *Bangladesh Demographic and Health Survey 2007,* Dhaka, Bangladesh and Calverton, Maryland, USA: National Institute of Population Research and Training, Mitra and Associates, and Macro International

—— (2005) *Bangladesh Demographic and Health Survey 2004,* Dhaka, Bangladesh and Calverton, Maryland, USA: National Institute of Population Research and Training, Mitra and Associates, and Macro International

Ravallion, Martin (2011) *Are we Really Assessing Development Impact?,* blog, Development Impact: News, Views, Methods, and Insights from the World of Impact Evaluation

—— (2009) 'Should the Randomistas Rule?', *The Economists' Voice* 6.2, Article 6

—— (2008) *Evaluation in the Practice of Development,* Policy Research Working Paper 4547, Washington, D.C.: World Bank

—— (2001) 'The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation', *World Bank Economic Review* 15.1: 115–40

Rawat, Kadiyala and McNamara (2010) 'The Impact of Food Assistance on Weight Gain and Disease Progression among HIV-infected Individuals Accessing AIDS Care and Treatment Services in Uganda', *BNC Public Health* 10: 316

Roetman, Eric (2011) *A Can of Worms? Implications of Rigorous Impact Evaluations for Development Agencies,* 3IE Working Paper #11, International Initiative for Impact Evaluation

Rozario, Santi (2004) *Building Solidarity Against Patriarchy,* Dhaka: CARE Bangladesh

Ruel, Marie; Menon, Purnima; Habicht, Jean-Pierre; Loechl, Cornelia; Bergeron, Gilles; Pelto, Gretel; Arimond, Mary; Maluccio, John; Michaud, Lesly and Hankebo, Bekele (2008) 'Age-based Preventive Targeting of Food Assistance and Behaviour Change and Communication for Reduction of Childhood Undernutrition in Haiti: A Cluster Randomized Trial', *Lancet* 371: 588–95

Rugh, Jim (2011) 'What's Involved in "Rigorous Impact Evaluation"? IOCE Proposes more Holistic Perspectives', presentation to NONIE Conference, Paris, 28 March 2011

Rugh, Jim; Steinke, Megan; Cousins, Brad and Bamberger, Michael (2010) *Summary of Discussion of the Alternative to the Statistical Counterfactual Think Tank,* AEA 2010 Annual Conference, San Antonio, Texas. www.realworldevaluation.org

Smith, Lisa C. and Haddad, Lawrence (2000) *Explaining Child Malnutrition in Developing Countries,* IFPRI Research Report 111, Washington, D.C.: International Food Policy Research Institute

Smith, Lisa C.; Ramakrishnan, Usha; Ndiaye, Aida; Haddad, Lawrence and Martorell. Reynaldo (2003) *The Importance of Women's Status for Child Nutrition in Developing Countries,* IFPRI Research Report 131, Washington, D.C.: International Food Policy Research Institute

SUN (Scaling Up Nutrition) (2010) *Scaling Up Nutrition: A Framework for Action,* based on a series of consultations hosted by the Center for Global Development, European Commission, International Congress of Nutrition, United Nations Standing Committee on Nutrition, USAID, UNICEF, WHO and the World Bank

Swindale, A.; Deitchler, Megan; Cogill, Bruce and Marchione, Thomas (2004) *The Impact of Title II Maternal and Child Health and Nutrition Programs on the Nutritional Status of Children,* Occasional Paper 4, USAID FANTA Project, Food and Nutrition Technical Assistance Project, Washington, D.C.: Academy for Educational Development

UNICEF (United Nations Children's Fund) (2009) *Tracking Progress on Child and Maternal Nutrition: A Survival and Development Priority,* New York: United Nations Children's Fund

—— (1998) *The State of the World's Children,* New York

Victora, Cesar G. (2009) 'Nutrition in Early Life: A Global Priority', *Lancet* 374.9696: 1123–5

Victora, Cesar G.; Adair, Linda; Fall, Caroline; Hallal, Pedro; Martorell, Reynaldo; Richter, Linda and Sachdev, Harshpal Singh (2008) 'Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital', *Lancet* 371.9609: 340–57

von Grebmer, Klaus; Ruel, Marie T.; Menon, Purnima; Nestorova, Bella; Olofinbiyi, Tolulope; Fritschel, Heidi and Yohannes, Yisehac (2010) *2010 Global Hunger Index: The Challenge of Hunger: Focus on the Crisis of Child Undernutrition,* Bonn, Washington, D.C., and Dublin: Deutsche Welthungerhilfe, International Food Policy Research Institute, and Concern

WFP, UNICEF, IPHN (2009) *Household Food Security and Nutrition Assessment in Bangladesh, November 2008–January 2009,* Dhaka: World Food Program, UNICEF and the Institute of Public Health Nutrition

White, Howard (2009) *Theory-based Impact Evaluation: Principles and Practice,* International Initiative for Impact Evaluation Working Paper 3, New Delhi: International Initiative for Impact Evaluation

—— (2008) 'Of Probits and Participation: The Use of Mixed Methods in Quantitative Impact Evaluation', *IDS Bulletin* 39.1: 98–109

White, Howard and Bamberger, Michael (2008) 'Introduction: Impact Evaluation in Official Development Agencies', *IDS Bulletin* 39.1: 1–11

World Bank (2010a) *Food Price Increases in South Asia: National Responses and Regional Dimensions,* South Asia Region, Sustainable Development Department, Agriculture and Rural Development Unit, Washington D.C.: World Bank

—— (2010b) *What can we Learn from Nutrition Impact Evaluations? Lessons from a Review of Interventions to Reduce Child Malnutrition in Developing Countries*, The Independent Evaluation Group, Washington, D.C.: World Bank

—— (2006) 'Repositioning Nutrition as Central to Development: A Strategy for Large-scale Action', *Directions in Development*, Washington, D.C.: The International Bank for Reconstruction and Development/The World Bank