Title
Evaluating population-level effects of water, sanitation, and hygiene interventions: methods and applications

Author
Benjamin-Chung, Jade De-Rong

Publication Date
2014

Evaluating population-level effects of water, sanitation, and hygiene interventions: methods and applications

by

Jade De-Rong Benjamin-Chung

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John M. Colford, Jr., Chair
Professor Alan E. Hubbard
Professor Kara L. Nelson

Spring 2014

Evaluating population-level effects of water, sanitation, and hygiene interventions: methods and applications

Abstract

Evaluating population-level effects of water, sanitation, and hygiene interventions: methods and applications

by

Jade De-Rong Benjamin-Chung

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor John M. Colford, Jr., Chair

Background: Scientists and development stakeholders argue that health interventions proven effective in randomized efficacy trials should be translated into large-scale programs to benefit public health. Substantive evidence supports the scale-up of numerous health interventions, such as water, sanitation, and deworming interventions, and since the establishment of the Millennium Development Goals (MDGs) the funding and motivation for such scale-up has grown. In the field of water and sanitation, numerous interventions have been demonstrated to be efficacious in the reduction of diarrhea and soil-transmitted helminth infection. However, scaling up these interventions to regional or national levels frequently presents implementation challenges, and systematically studying the reasons for scale-up success or failure is essential to refine and sustain public health programs. Another important feature of scaling up interventions is determining how best to integrate interventions at scale and whether intervention delivery should be focused at the individual, household, or community level. Population attributable fraction (PAF) parameters and a new class of parameters which build upon the PAF can be used to estimate the effect of large-scale programs on population health. Evaluation of interventions at scale poses unique questions, and epidemiologic designs and analyses need to be tailored to answer these particular questions. Modern approaches to PAF estimation allow for parameter definition to be tailored one's particular research question and are well suited to the evaluation of population-level effects of large-scale health interventions.

Methods: In this dissertation, I illustrate and apply methods to evaluate population-level effects of water, sanitation, and hygiene interventions. I specifically focus on methods for and applications with observational, cross-sectional data, and I discuss generalizations to other study designs. In the first chapter, I quantify the association between deworming, improved sanitation, and hygiene interventions and soil-transmitted helminths in a population in rural Bangladesh. I assess the potential for interactions between these interventions and explore associations at both the individual and village level. In the second chapter I assess the quality

of implementation of a large-scale water, sanitation, and hygiene intervention implemented by UNICEF and the Government of Bangladesh in rural Bangladesh. It was found that this intervention did not meet most of its health and behavior targets in an interim evaluation. To help understand why, I envision a scenario in which implementation had been better in all areas, and I estimate how much outcomes may have changed under this scenario compared to the outcomes that were observed. In the third chapter, I discuss parameters appropriate for estimating population-level effects of health interventions. Specifically, I describe the estimation of the PAF and two modern parameters which build upon the PAF: the population intervention model and stochastic intervention model parameters. I provide a didactic description of the estimation of these parameters.

Significance: This dissertation illustrates the use of rigorous methods to systematically evaluate the effect of individual and combined interventions at scale. Rigorous assessment of water, sanitation, and hygiene interventions is difficult, even for small-scale interventions, and very few large-scale WASH interventions have been evaluated rigorously. The parameters I illustrated and estimated in this dissertation have broad applicability to similar assessments of other large-scale public health programs. My findings contribute to the growing empirical evidence base describing best practices for and barriers to delivering interventions at scale. This evidence may contribute to improvements in design, delivery, and prioritization of interventions which in turn could increase the health impact of such interventions when delivered at scale.

For my parents. I am so grateful for their continued support and encouragement.

Contents

List of Figures

List of Tables

Acknowledgments

I would first like to acknowledge my advisor, Jack Colford, for his consisent mentoring and support throughout my graduate training. Jack's accessibility, insightfulness, and strategic thinking have been critical to my development as an epidemiologist. He has and continues to provide me with numerous opportunities to build my skillset, grow my professional network, and challenge myself. He is truly a one of a kind mentor.

Thank you to my other committee members, Alan Hubbard and Kara Nelson, for their valuable input on this dissertation. I am very grateful to Ben Arnold for his generous advice and mentorship throughout my time at UC Berkeley. I thank Steve Luby for his mentorship, humor, and thoughtfulness in providing input on my scientific work and the evolution of my career. Thank you to David Addiss for his consistently thoughtful input in my scientific work. Thank you to Art Reingold and Ira Tager for guiding me in my early studies of epidemiology and encouraging me down the path towards a PhD.

Thanks to my collaborators at the International Centre for Diarrhoeal Disease Research, Bangladesh, whose efforts were critical to collecting the data used in this study, conducting laboratory analyses, and familiarizing me with research and life in Bangladesh: Leanne Unicomb, Sonia Sultana, Amal Halder, Tarique Huda, Mohammed Ali Ahsan, Tania Bulbul, Yushuf Sharker, Arifa Nazneen, Abul Kasham Shoab, Mohammed Abdur Raquib, Bimal Kumar Das, Rashidul Haque, and Abdullah Siddique, among many others.

Thank you to the University of California Global Health for partially funding the field and laboratory work in this dissertation through a Student Fellowship. I would like to thank Davide Ianeillo and Andrew Majewski for providing training to our laboratory team on mini-FLOTAC and the University of Naples for providing the mini-FLOTAC devices.

Finally, thank you to my supportive friends, particularly those who helped me survive my biostatistics coursework: Jeremy Coyle, Marco Carone, Anna Decker, Deb Karasek, Hannah Leslie, Morgan Levy, and Robin Mejia.

Chapter 1

Introduction

1.1 Motivation

Numerous health interventions have been shown to be efficacious at a small scale and in randomized trials [1–3]. The establishment of the Millennium Development Goals (MDGs) has spurred an increase in funding and motivation for the scale-up of such interventions to benefit public health [4]. In the field of water, sanitation, and hygiene (WASH), numerous interventions have been demonstrated efficacious to reduce diarrhea and soil-transmitted helminth infection on small scales [11–24], and the scale up of these interventions is increasing. For example, from 2007-2012, UNICEF and Government of Bangladesh delivered a WASH program to 20.4 million people in Bangladesh [25]. Through the Water and Sanitation Program, the World Bank has delivered and evaluated the impact of WASH interventions to tens of thousands of people in India, Indonesia, Peru, Vietnam, and other countries [26–29]. The field of neglected tropical diseases (NTDs) has increasingly explored the scale-up of WASH interventions to prevent NTDs, and there has been a call for greater integration of large-scale NTD programs, such as mass drug administration, and large-scale WASH interventions [30].

Scaling up health interventions frequently prompts questions about how interventions should be delivered and integrated [31–34]. Epidemiologic designs and analyses need to be tailored to answer these particular questions. First, questions related to the quality of implementation arise: when compliance is poor or delivery is incomplete at a large scale, it can be difficult to determine whether an observed lack of public health impact reflects a poor intervention design or an intervention that could not be implemented well at scale. The reasons for poor compliance or incomplete intervention delivery at scale frequently differ from the reasons in small-scale settings. Few large-scale interventions, particularly in the WASH sector, have been evaluated rigorously, and the evaluations that have been done of scaled up WASH interventions found no effect on access to improved sanitation and mixed results related to handwashing and diarrhea prevalence [27–29]. Second, when evaluating interventions at scale, it is frequently the case that the population also receives other large-scale interventions concurrently. For example, many WASH programs aim to reduce not only diarrhea but also soil transmitted helminth infection. Populations targeted by WASH programs often also receive school-based mass administration of deworming, and it is possible that when deworming and WASH interventions are delivered concurrently, they interact synergistically, yielding greater improvements in health than would be expected. Because reinfection with soil-transmitted helminths typically occurs rapidly following deworming [35], in order to sustainably reduce the burden at the population-level, provision of both deworming and WASH intervenions may be needed. The existing literature has largely assessed these two sets of interventions separately, but careful assessment of the potential interaction between them is critical to planning scale up efforts for either intervention.

In evaluations of community-based public health interventions, randomized trials remain the gold standard in epidemiology, and recently their use has grown in related fields, such as economics [36]. The chief advantage of using trials is their high internal validity, however, they can have limited generalizability and utility for important research questions in public health, particularly when one's aim is to assess the population-level effectiveness of interventions. In addition, it is typically neither feasible nor ethical to randomize when evaluating the effectiveness of a large-scale interventions known to be efficacious in ideal settings [37]. For these and other reasons, observational designs, while subject to many pitfalls of their own, should not be overlooked. Analyses of observational data are often criticized because their inference relies upon the statistical model rather than the study design [38]. Another critique is that the choice of which quantity to estimate is often determined by the statistical model used instead of by the research question [39]. Statistical approaches can never remedy a poorly designed study. However, this dissertation demonstrates how to carefully define parameters to estimate with observational data that are tailored to the specific research question. These approaches are broadly applicable but are particularly useful when evaluating large-scale interventions' effects on population-level health outcomes.

1.2 Specific aims

In this dissertation, I illustrate and apply methods to evaluate population-level effects of water, sanitation, and hygiene interventions. I specifically focus on methods for and applications with observational, cross-sectional data, and I discuss generalizations to other study designs. My specific aims are as follows:

1. To explore potential interactions between deworming, sanitation, and hygiene interventions (Chapter 2).

2. To estimate the extent to which hygiene behavior and conditions may have improved if the SHEWA-B program had been better implemented (Chapter 3).

3. To illustrate how to estimate and interpret the population attributable fraction, the population intervention model parameter, and the stochastic intervention model parameter using simulated and empirical datasets (Chapter 4).

My first two aims analyze empirical data from an evaluation of SHEWA-B, the abovementioned large-scale WASH intervention implemented by UNICEF and the Government of Bangladesh. I collaborated with the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B), which led the evaluation of SHEWA-B. My first aim utilizes data from the evaluation but does not specifically evaluate the SHEWA-B program, whereas the second aim does. The third aim has a methodologic focus, but it uses the SHEWA-B evaluation data from the second aim to illustrate estimation in an empirical dataset.

Bibliography

[1] Bryce J, Black RE, Walker N, Bhutta ZA, Lawn JE, and Steketee RW. Can the world afford to save the lives of 6 million children each year? *The Lancet* 365:9478 (2005), pp. 2193–2200.

[2] Jha P, Mills A, Hanson K, Kumaranayake L, Conteh L, Kurowski C, Nguyen SN, Cruz VO, Ranson K, Vaz LM, et al. Improving the health of the global poor. *Science* 295:5562 (2002), pp. 2036–2039.

[3] Victora CG, Hanson K, Bryce J, and Vaughan JP. Achieving universal coverage with health interventions. *The Lancet* 364:9444 (2004), pp. 1541–1548.

[4] Sachs JD. Health in the developing world: achieving the Millennium Development Goals. *Bulletin of the World Health Organization* 82:12 (2004), pp. 947–949.

[5] Bloom G and Ainsworth P. *Beyond scaling up: pathways to universal access to health services*. STEPS Working Paper 40. STEPS Centre, 2010.

[6] Tangcharoensathien V and Patcharanarumol W. Commentary: Global health initiatives: opportunities or challenges? *Health Policy and Planning* 25:2 (2010), pp. 101–103.

[7] Lu C, Schneider MT, Gubbins P, Leach-Kemon K, Jamison D, and Murray CJ. Public financing of health in developing countries: a cross-national systematic analysis. *The Lancet* 375:9723 (2010), pp. 1375–1387.

[8] *World Development Report 1993: Investing in Health*. World Bank and Oxford University Press, 1993.

[9] *Scaling Up: From Vision to Large-Scale Change*. Management Systems International, 2012. URL: `http://www.msiworldwide.com/wp-content/uploads/Scaling-Up-Framework.pdf`.

[10] Duflo E. *Scaling Up and Evaluation*. 2004. URL: `http://economics.mit.edu/files/766`.

[11] Clasen TF, Bostoen K, Schmidt WP, Boisson S, Fung I, Jenkins MW, Scott B, Sugden S, and Cairncross S. Interventions to improve disposal of human excreta for preventing diarrhoea. *Cochrane Database of Systematic Reviews* 6 (2010).

[12] Clasen T, Schmidt WP, Rabie T, Roberts I, and Cairncross S. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *BMJ* 334:782 (2007).

[13] Curtis V and Cairncross S. Effect of washing hands with soap on diarrhoea risk in the community: a systematic review. *The Lancet Infectious Diseases* 3:5 (2003), pp. 275–281.

[14] Daniels DL, Cousens SN, Makoae L, and Feachem RG. A case-control study of the impact of improved sanitation on diarrhoea morbidity in Lesotho. *Bulletin of the World Health Organization* 68:4 (1990), pp. 455–463.

[15] Doocy S and Burnham G. Point-of-use water treatment and diarrhoea reduction in the emergency context: an effectiveness trial in Liberia. *Tropical Medicine & International Health* 11:10 (2006), pp. 1542–1552.

[16] Ejemot-Nwadiaro R, Ehiri J, Meremikwu M, and Critchley J. Hand washing for preventing diarrhoea. *Cochrane Database of Systematic Reviews* 1 (2008).

[17] Fewtrell L, Kaufmann RB, Kay D, Enanoria W, Haller L, and Colford Jr JM. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *The Lancet Infectious Diseases* 5:1 (2005), pp. 42–52.

[18] Luby SP, Agboatwalla M, Feikin DR, Painter J, Billhimer W, Altaf A, and Hoekstra RM. Effect of handwashing on child health: a randomised controlled trial. *The Lancet* 366:9481 (2005), pp. 225–233.

[19] Rabie T and Curtis V. Handwashing and risk of respiratory infections: a quantitative systematic review. *Tropical Medicine & International Health* 11:3 (2006), pp. 258–267.

[20] Roberts L, Chartier Y, Chartier O, Malenga G, Toole M, and Rodka H. Keeping clean water clean in a Malawi refugee camp: a randomized intervention trial. *Bulletin of the World Health Organization* 79:4 (2001), pp. 280–287.

[21] Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, Jha P, Campbell H, Walker CF, Cibulskis R, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *The Lancet* 375:9730 (2010), pp. 1969–1987.

[22] Prüss-Üstün A, Bos R, Gore F, and Bartram J. *Safer water, better health: costs, benefits and sustainability of interventions to protect and promote health.* World Health Organization, 2008.

[23] Ziegelbauer K, Speich B, Mäusezahl D, Bos R, Keiser J, and Utzinger J. Effect of sanitation on soil-transmitted helminth infection: systematic review and meta-analysis. *PLoS Medicine* 9:1 (2012), e1001162.

[24] Strunz EC, Addiss DG, Stocks ME, Ogden S, Utzinger J, and Freeman MC. Water, Sanitation, Hygiene, and Soil-Transmitted Helminth Infection: A Systematic Review and Meta-Analysis. *PLoS Medicine* 11:3 (2014), e1001620.

[25] Huda TMN, Unicomb L, Johnston RB, Halder AK, Yushuf Sharker MA, and Luby SP. Interim evaluation of a large scale sanitation, hygiene and water improvement programme on childhood diarrhea and respiratory disease in rural Bangladesh. *Social Science & Medicine* 75:4 (2012), pp. 604–611.

[26] Patil SR, Arnold BF, Salvatore A, Briceno B, Colford Jr JM, and Gertler PJ. A randomized, controlled study of a rural sanitation behavior change program in Madhya Pradesh, India. *World Bank Policy Research Working Paper* 6702 (2013).

[27] Cameron L, Shah M, and Olivia S. *Impact evaluation of a large-scale rural sanitation project in Indonesia*. World Bank, Washington, DC, 2013.

[28] Galiani S, Gertler P, and Orsola-Vidal A. *Promoting Handwashing Behavior in Peru: The Effect of Large-Scale Mass-Media and Community Level Interventions. SSRN Scholarly Paper*. 2012. URL: `http://papers.ssrn.com/abstract=2170640`.

[29] Chase C and Do QT. *Scaling Up Handwashing Behavior: Findings from the Impact Evaluation Baseline Survey in Vietnam*. World Bank, Washington, DC, 2010.

[30] Freeman MC, Ogden S, Jacobson J, Abbott D, Addiss DG, Amnie AG, Beckwith C, Cairncross S, Callejas R, Colford Jr JM, et al. Integration of water, sanitation, and hygiene for the prevention and control of neglected tropical diseases: a rationale for inter-sectoral collaboration. *PLoS Neglected Tropical Diseases* 7:9 (2013), e2439.

[31] Madon T, Hofman K, Kupfer L, and Glass R. Implementation Science. *Science* 318:5857 (2007), pp. 1728–1729.

[32] McCannon CJ, Berwick DM, and Massoud MR. The science of large-scale change in global health. *JAMA* 298:16 (2007), pp. 1937–1939.

[33] Mangham LJ and Hanson K. Scaling up in international health: what are the key issues? *Health Policy and Planning* 25:2 (2010), pp. 85–96.

[34] Sanders D and Haines A. Implementation research is needed to achieve international health goals. *PLoS Medicine* 3:6 (2006), e186.

[35] Jia TW, Melville S, Utzinger J, King CH, and Zhou XN. Soil-transmitted helminth reinfection after drug treatment: a systematic review and meta-analysis. *PLoS Neglected Tropical Diseases* 6:5 (2012).

[36] Duflo E, Glennerster R, and Kremer M. "Using randomization in development economics research: a toolkit". *Handbook of Development Economics*. Vol. 4. Elsevier, B.V., 2008. Chap. 61.

[37] Farmer P. *Clinical trials and global health equity*. http://globalhealth.thelancet.com/2013/07/08/clinical-trials-and-global-health-equity. 2013.

[38] Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine* 26:1 (2007), pp. 20–36.

[39] van der Laan MJ and Rose S. *Targeted learning: causal inference for observational and experimental data*. Springer, 2011.

Chapter 2

The interaction of deworming, improved sanitation, and household flooring and soil-transmitted helminth infection in rural Bangladesh

2.1 Background

The World Health Organization recommends mass drug administration (MDA) with anthelmintics as well as improved sanitation and personal hygiene to reduce the prevalence and transmission of soil-transmitted helminths (STH) in endemic countries [1]. For example, in Bangladesh where I conducted the present study, the Bangladesh Ministry of Health and Family Welfare has implemented national MDA of mebendazole in schools twice annually since 2008, and the program currently targets children aged 5 to 14 years. The Bangladesh Expanded Program on Immunization deworms pre-school children in Bangladesh nationally. In 2005, prior to MDA, an estimated 80% of Bangladeshi school-age children were infected with STH [2]. Since the initiation of MDA in Bangladesh, to my knowledge there have not been any systematic surveys of STH prevalence. Updated prevalence estimates will inform government officials and other health providers in Bangladesh about whether to continue MDA, modify it, or provide additional complementary interventions to control STH infections.

Despite the high efficacy of anthelmintics to reduce infection prevalence in the short term, a meta-analysis estimated that within six months, 68% (95% CI 60-76%) of those treated become reinfected with *Ascaris*, 67% (95% CI 42-100%) with *Trichuris*, and 55% (95% CI 34-87%) with hookworm [3]. A large body of evidence, largely from observational and cross-sectional studies, suggests that improved sanitation can reduce the risk infection or reinfection with soil-transmitted helminths (STH) [4–7]. In addition to sanitation, there is strong biological plausibility to support the provision of finished floors (i.e., cement or wood floors) as an intervention to decrease the risk of STH infection. STH eggs must be deposited in the soil to reach their infective stages; provision of finished floors to households with earthen floors thus removes the majority infective stages from the indoor living environment, reducing the probability of transmission. While larvae and ova may still be present on surfaces in households with finished floors, their survival time is likely to be shorter. Few studies have systematically explored whether finished flooring reduces the risk of STH infection; three studies identified an association between living in a household with an earthen floor and increased risk of STH infection, however these studies did not adjust for household wealth, a potentially strong confounder of this association [8–10].

There has been a call to consider the joint effects of anthelminthics and water, sanitation, and hygiene together in order to identify more sustainable methods of reducing STH infection and transmission [6, 11], yet few studies have done so [12–15]. Only one study has explicitly explored finished floors as an intervention to reduce parasite infection, and it did not measure STH infection [16]. Furthermore, no studies have formally explored whether water, sanitation, and hygiene interventions and MDA interventions could yield greater risk reductions when delivered in combination (i.e., whether there is evidence of synergy) [17]. Evidence of a synergistic interaction between these interventions would motivate the development and delivery of combined interventions to more sustainably reduce the incidence and transmission of STH. Among practitioners and policymakers, control of STH and other neglected tropical diseases has largely been a separate enterprise from control of enteric pathogens through water, sanitation, and hygiene interventions [11]. This is the case in Bangladesh, where the government administers MDA and large international non-governmental agencies, such as UNICEF, BRAC, and the Grameen Bank, deploy the majority of sanitation and hygiene interventions [18, 19].

Transmission models predict that increasing deworming and sanitation coverage at the community level would reduce the prevalence of infection [20–22]. Clustering of infection in communities may reflect differences in susceptibility and immunological response due to genetics, as well as household-level heterogeneity in exposure. Even though numerous studies have described clustering of STH infection at the household level and high aggregation of STH within communities [10, 20, 21, 23–29], extensive deworming, sanitation, or finished flooring coverage would in theory result in reduced prevalence and transmission of STH because 1) individuals are likely to be exposed not only in their homes but also in other areas of a village, for instance, while at school or work, and 2) empirical evidence and modeling studies have described herd effects of deworming [22, 30, 31]. Specifically, they have found a substantial decrease in STH prevalence following provision of school-based deworming with over 90% coverage not only in school-age children but also younger children and adults [30, 31]. One study in Kenya found evidence of decreased STH infection among children who attended schools that did not offer school-based deworming but were near those that did [32]. Understanding the extent to which cluster-level coverage of exposures is associated with STH infection would aid in the targeting of future interventions.

The objectives of this study were to: 1) estimate the prevalence of STH infection among children and women of childbearing age in rural Bangladesh, 2) estimate associations with deworming, hygienic latrines, and finished floors and STH infection, 3) explore potential interactions between these exposures, and 4) estimate associations between cluster-level exposures and cluster-level STH prevalence.

2.2 Methods

Study population and sample

This study was conducted as part of a larger study evaluating the Sanitation Hygiene Education and Water Supply in Bangladesh (SHEWA-B) program, which was implemented by UNICEF and the Government of Bangladesh. This particular study was not focused on the evaluation of SHEWA-B itself, which is being conducted in part by the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B), but it does leverage the cross-sectional survey collected in 68 sub-districts and 19 districts of rural Bangladesh as part of the endline evaluation of SHEWA-B in 2012. I conducted this study in the 100 intervention and control village clusters selected for the SHEWA-B endline evaluation. The intervention, selection of control areas, and sampling of clusters in the intervention and control areas for the SHEWA-B evaluation have been described elsewhere [33].

In each selected village cluster, the field team identified the center point of the village and the nearest eligible household. The team skipped the nearest two eligible households and enrolled 18 households per cluster. Households were eligible if a child under five years resided there at the time of data collection. In each cluster, the aim was to collect data from six people within each of the following age and sex categories: children 1-4 years, children 5-14 years, and women 15-49 years. Within each cluster the field team determined the number of eligible people in each age group. If there were less than six eligible individuals available in a particular age group, the team enrolled an additional person from the next cluster. If there were multiple persons available in a particular age group, the team chose the youngest individual within the age group in the cluster. The field team collected stool samples and initial questionnaires about socio-demographic information and deworming history in October 2012. In December 2012, the field team returned to households where stool samples were collected and administered a questionnaire to measure household and environmental exposures. These exposures were ascertained following stool sample collection due to field logistics constraints and the need to complete stool collection prior to national MDA in early November 2012.

Stool specimen collection and analysis

Field workers provided households with plastic sheets and stool collection tubes in which to collect stool samples and returned in 24 hours to collect the samples. Upon retrieving the sample, field workers weighed 1g of stool and placed it in 20 ml of 4% sodium acetate-acetic acid-formalin and thoroughly homogenized the stool in the formalin. The maximum time between defecation and stool processing was 12 hours. Samples were transported to Dhaka, Bangladesh for laboratory analysis. Helminth ova were detected using mini-FLOTAC, a copromicroscopic diagnostic technique appropriate for preserved stool [34, 35]. Laboratory staff centrifuged samples at 1500 RPM for 3 minutes and then discarded the supernatant and suspended the sedimented stool in 20 ml of flotation solution 2 (saturated sodium chloride), mixed the contents, and filled each of the two chambers of the mini-FLOTAC device with 1ml of the mixed sample. Staff recorded the number of eggs of *Ascaris* lumbricoides, hookworm, *Trichuris trichiura*, and *Enterobius vermicularis* in each chamber. For each helminth, I averaged the number of eggs in each chamber and multiplied the number by a factor of 10 to quantify the number of eggs per gram of stool.

Outcome and exposure definitions

Outcomes included presence of any helminth ova and intensity of helminth infection. Moderate/high intensity infections were defined as $\geq$5,000 eggs/gram for *Ascaris*, $\geq$1,000 eggs/gram for *Trichuris*, and $\geq$2,000 eggs/gram for hookworm [36]. Exposures include access to a hygienic latrine, household flooring material (earth/bamboo or cement/wood), and self-reported deworming in the last six months. I defined hygienic latrines as flush latrines connected to piped sewer system, to a septic tank, or off-set pit latrine, pit latrine with slab and functional water seal, pit latrine with slab, lid and no water seal, or a composting latrine. I defined unhygienic latrines as those that fail to effectively separate feces from the environment: flush latrines connected to canal or ditch, pit latrines without a slab, pit latrines with a slab, no or broken water seal or a hanging latrine. This definition was developed by the ICDDR,B and is intended to more accurately categorize latrines that isolate feces from the environment in the Bangladeshi context than the commonly used WHO Joint Monitoring Programme (JMP) definition [37] (see Table A.1). Specifically, hygienic latrines require a water seal or a lid on a pit to effectively separate collected feces from the environment, and I do not consider sharing status of a latrine. For each person who provided a stool sample, the field team asked the respondent whether that person was dewormed in the last six months and if so, approximately how many weeks or months ago the person was dewormed. They also asked whether deworming was received as part of a campaign and the source of deworming (e.g. clinic, school).

I calculated the cluster-level deworming coverage as the percentage of respondents in the sample who reported being dewormed in the prior six months in a given cluster. To estimate cluster-level sanitation and finished floor coverage, I calculated the percentage of respondents with each exposure in a cluster.

I identified potential confounders using directed acyclic graphs [38] (see Figure A.1), and I controlled for these potential confoudners in statistical models used. These included age, sex, cluster-level wealth, household wealth, and mother's education level. The field team collected information about the presence of household assets (e.g. refrigerator, mobile phone) and used principal components analysis to develop an index of household wealth [39] (see Table A.2). Households in the lowest three quintiles of the first principal component were classified as lower household wealth and those in the highest two quintiles were classified as higher household wealth. Cluster-level wealth was calculated as the percentage of households in the fourth and fifth quintiles of household wealth.

Sample size

Since estimates of STH prevalence for the age and sex groups of interest were not available for the study areas, to be conservative, I assumed the prevalence of all helminths was 50%. I assumed a design effect of 2.6, which is based on intra-class correlation coefficients estimated at the village level for *Ascaris*, hookworm, and *Trichuris* infection in children under 5 years from a study of sanitation Tamil Nadu, India in 2008 since information from Bangladesh was not available [40]. My calculations assumed a fixed sample size of 1,700 (100 village clusters $\times$ 17 individuals per cluster). Under these assumptions, the precision associated with an estimate of prevalence of 50% is $\pm$4% .

Statistical analysis

I calculated pooled and age- and sex-specific prevalence by type of helminth. To examine the association between prevalence and cluster-level variables, I produced scatter plots of the observed variables and used smoothed locally weighted scatter plot smoothing (i.e. LOWESS) with normal-based 95% confidence bands to explore patterns in each scatter plot [41]. I also estimated the intraclass correlation coefficient for each STH infection within each cluster using a one-way analysis of variance.

To estimate adjusted prevalence ratios I used modified Poisson (i.e. log linear) regression [42]. The model adjusted for the potential confounders defined above. I also estimated the prevalence ratio using a semiparametric procedure with a data-adaptive machine learning approach [43]. The learners included generalized linear models, Bayesian main-terms logistic regression, lasso and elastic-net regularized GLM, generalized additive models, and stepwise regression with only main effect terms based on the Akaike Information Criterion. Point estimates from the more complex semiparametric estimator were similar to those from the modified Poisson regression, so I present only the regression results. I estimated robust standard errors clustered at the village level to account for potential within-village outcome correlation. I excluded from the analysis individuals with missing outcomes, which assumes that they were missing completely at random.

Standard statistical models for binary outcomes predict outcomes on the multiplicative scale, and accordingly, interaction is often assessed on the multiplicative scale. However, there is some consensus that in a public health context, interactions are better assessed on the additive scale [17, 44, 45]. This is particularly the case when one's aim is to assess synergy, or departures from additivity of associations, rather than statistical interaction, or the interdependence of two risk factors within a particular statistical model [44]. Since my aim is to understand whether synergy is present, I estimated the relative excess risk due to interaction (RERI), a measure of additive interaction that can be calculated from multiplicative models [46]. Since I expected associations to be protective, I recoded variables prior to RERI calculation so that the stratum with the PR furthest from the null was reassigned as the reference group [47]. I also estimated the ratio of prevalence ratios, which assess interaction on the multiplicative scale [48]. I report prevalence ratios and their accompanying confidence intervals within strata of deworming and strata of hygienic latrine access and finished flooring coverage. Because data were clustered at the village level, I used the bootstrap and resampled clusters to estimate 95% confidence intervals. I did not estimate confidence intervals for any point estimates for which there were strata with fewer than 5 units. Analyses were conducted in Stata version 12 and in **R** version 3.0.2.

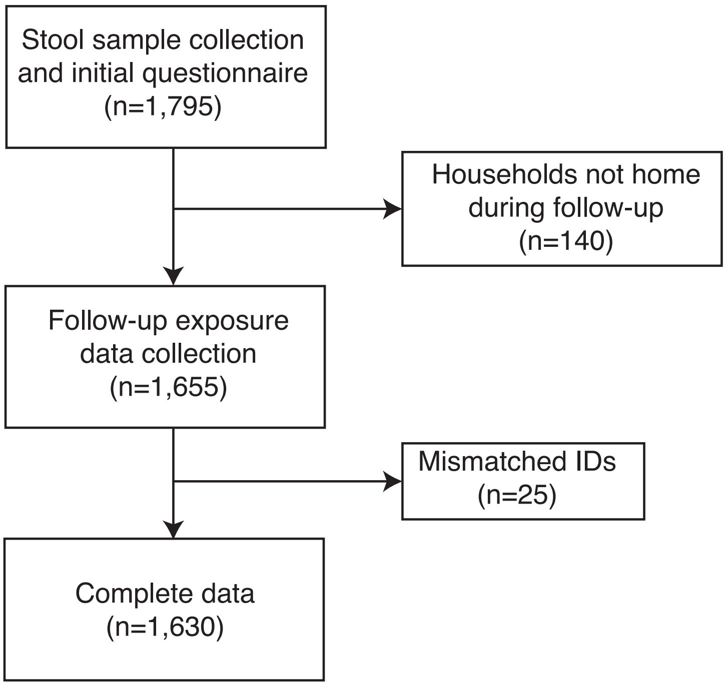
2.3 Results

The field team collected stool samples and initial questionnaires from 1,795 individuals in October 2012. In December 2012 they collected 1,655 surveys with exposure information, including household access to an improved latrine and household floor material as well as demographic information; 140 of the households visited in October were not home in December. Another 25 households had mismatched IDs. The complete dataset with both exposure and outcome information used in this analysis contained 1,630 observations (Figure 2.1).

Less than half (40%, n=656) of mothers had at least a primary education. Household size ranged from 2 to 18, with a mean of 5.5, and household size was not associated with STH infection. About a third of households (32%, n=527) had access to a hygienic latrine, and 13% (n=207) lived in households with finished floors. Respondents reported that 49% of children 1-4 years, 52% of children 5-14 years, and 21% of women of childbearing age took deworming medication in the prior six months. Slightly under half (47%) of school-age children were reported to have been dewormed at school, and for other age groups, the predominant source of deworming was the home or a source in the village, such as a local pharmacy.

Approximately one third (32%) of individuals sampled had an STH infection, and 9% had multiple infections (Table 2.1). Single and multiple infections were most common among school-age children. Across all age groups, *Trichuris* was the most prevalent infection, infecting 17% of children 1-4 years, 28% of children 5-14 years, and 18% of women of child-bearing age. Across all ages, 2.7% of respondents were infected with *Enterobius*, and 1.6% of school-aged children were infected. For all helminths and age groups, less than 2% had moderate/heavy intensity infections; there were no individuals with moderate/heavy intensity hookworm infections.

Figure 2.1: Data collected



I found protective associations with individual exposures of interest for *Ascaris* and hookworm prevalence, but associations were close to null for *Trichuris* prevalence (Table 2.2). For *Ascaris*, the adjusted prevalence was 0.60-fold lower among those who were dewormed compared to among those who were not and 0.52-fold lower among those with finished floors compared to those with unfinished floors, and the associations were statistically significant. The adjusted *Ascaris* prevalence was 0.88-fold lower among those with access to a hygienic latrine compared to among those without, but the association was not statistically significant (Table 2.2; Table A.1). For hookworm the adjusted prevalence ratio for the association with deworming was 0.91, and it was not statistically significant. There was a stronger protective association between hookworm and access to a hygienic latrine (aPR=0.75) and finished floors (aPR=0.44), but these findings also were not statistically significant. For *Trichuris*, the associations were close to the null and not statistically significant: the aPR was 1.02 for deworming, 1.00 for access to a hygienic latrine, and 1.01 for finished floors. Household size was not associated with STH infection.

Table 2.1: Helminth infection by respondent age, organism, and infection intensity

	Child 1-4 years (n=549)	Child 5-14 years (n=549)	Women 15-49 years (n=532)	All (n=1630)
Female (%)	47.7	50.3	100.0	65.6
Dewormed in last six months (%)	49.3	52.4	20.9	41.0
Mean months since deworming	2.8	3.2	2.3	2.8
Source of deworming				
Home/village	68.7	37.4	76.6	56.5
Health clinic	26.9	15.7	22.5	21.4
School	3.4	46.9	0.0	21.5
Other	1.1	0.0	0.9	0.6
Any infection*	25.7	40.1	30.3	32.0
Multiple infections*	7.8	12.4	7.5	9.3
Ascaris				
Prevalence (%)	12.9	14.4	11.8	13.1
Mean eggs per gram	318	279	387	287
No infection (%)	87.1	85.6	88.2	86.9
Light infection (%)	11.8	12.6	10.7	11.7
Moderate/heavy infection (%)	1.1	1.8	1.1	1.3
Hookworm				
Prevalence (%)	2.6	7.7	6.4	5.5
Mean eggs per gram	8	2	12	10
No infection (%)	97.4	92.3	93.6	94.5
Light infection (%)	2.6	7.7	6.4	5.5
Moderate/heavy infection (%)	0.0	0.0	0.0	0.0
Trichuris				
Prevalence (%)	17.1	27.5	18.2	21.0
Mean eggs per gram	43	32	74	22
No infection (%)	82.9	72.5	81.8	79.0
Light infection (%)	16.8	26.4	18.0	20.4
Moderate/heavy infection (%)	0.4	1.1	0.2	0.6
Enterobius				
Prevalence (%)	0.01	0.05	0.02	0.03
Mean eggs per gram	4	0	10	2
No infection (%)	–	–	–	–
Light infection (%)	–	–	–	–
Moderate/heavy infection (%)	–	–	–	–

*Includes *Enterobius* infections

Table 2.2: Prevalence ratios for deworming, hygienic latrine access, and finished floors

	n	*Ascaris* PR (95% CI)	Hookworm PR (95% CI)	*Trichuris* PR (95% CI)
Unadjusted prevalence ratios				
Deworming	1622	0.60 (0.46,0.80)	0.79 (0.52,1.21)	1.02 (0.84,1.24)
Access to hygienic latrine	1629	0.78 (0.59,1.04)	0.60 (0.37,0.97)	0.93 (0.75,1.14)
Finished floor	1630	0.45 (0.26,0.77)	0.32 (0.12,0.86)	0.88 (0.66,1.19)
Adjusted prevalence ratios*				
Deworming	1605	0.60 (0.45,0.79)	0.91 (0.60,1.39)	1.02 (0.84,1.24)
Access to hygienic latrine	1612	0.88 (0.65,1.19)	0.75 (0.44,1.25)	1.00 (0.80,1.24)
Finished floor	1613	0.52 (0.30,0.90)	0.44 (0.15,1.29)	1.01 (0.73,1.39)

*PRs estimated using log binomial regression and adjusted for age, sex, sub-district, household wealth, cluster-level wealth, and mother's education level

To explore potential interactions, I plotted the prevalence under each individual and joint exposure level (Figure 2.2) and estimated aPRs for separate and combined exposures and measures of interaction on the additive scale (Relative Excess Risk due to Interaction – RERI) and multiplicative scale (Ratio of Prevalence Ratios – RPR) (Tables 2.3 and 2.4). Tables A.3 and A.4 display these results in accordance with presentation format recommended by Knol and VanderWeele [48]. When the RERI is equal to zero, there is evidence of no interaction on the additive scale. An RERI less than zero indicates that the combined association is less than the sum of the individual associations (antagonistic or subadditive), and when it is greater than zero, the combined association is greater than the sum of the individual associations (synergistic) [44, 47]. When the exposures of interest are associated with only a lower or higher prevalence (i.e., they are monotonic), an RERI>0 indicates a synergistic interaction between exposures [49, 50]. If the exposures are not monotonic then the RERI must be greater than 1 for synergistic interaction to be present [49, 50].

Table 2.3 assesses potential interactions between access to a hygienic latrine and deworming. For *Ascaris*, the aPR was 0.63 for deworming alone, 0.97 for hygienic latrine access alone, and 0.44 both. The RERI was -0.43, indicating that the joint aPR was closer to the null than the sum of the individual aPRs; however, the RERI was not statistically significant (95% CI -2.40, 0.48). The RPR was 0.77, suggesting interaction on the multiplicative scale, but it was not statistically significant (95% CI 0.35, 1.53). I found a similar pattern for hookworm: the aPR was 0.82 for deworming alone, 0.55 for access to hygienic latrines alone, and 0.81 for both. The RERI was 0.45 but was not statistically significant (95% CI -0.88, 1.15), suggesting a possible synergistic interaction between deworming and hygienic latrine access for hookworm. For *Trichuris*, the aPR was 1.20 for hygienic latrine access alone, 1.20 for deworming alone, and 0.87 for both. The RERI for the potential interaction between hygienic latrine access and deworming for *Trichuris* was -0.65 (95% CI -1.34, -0.12), indicating a statistically significant interaction on the additive scale. While the value of the

RERI was negative and thus does not indicate synergy, from a public health perspective, the finding that joint exposures were protective and individual exposures were not supports the exploration of combined interventions to reduce *Trichuris* prevalence.

Table 2.4 explores interaction between living in a household with finished floors and deworming. For *Ascaris*, the aPR was 0.60 for deworming alone, 0.54 for finished flooring alone, and 0.29 for both. The RERI was 0.51, indicating that the aPR for both exposures jointly exceeded the sum of the individual aPRs and indicating synergistic interaction under the monotonicity assumption; however, the results were not statistically significant (95% CI -3.67, 1.91). The aPRs for hookworm were 0.94 for deworming alone, 0.50 for finished floors alone, and 0.26 for both. The RERI for hookworm was negative, and the confidence intervals for the RERI and RPR for hookworm were not estimated due to sparse data. For *Trichuris*, a similar pattern to that for the interaction between deworming and access to hygienic latrines was observed with finished floors: the aPRs were 1.05 for deworming alone, 1.10 for finished floors, and 0.96 for both. The RERI was -0.23 (95% CI -1.59, 0.37). While many of the findings in tables 3 and 4 were not statistically significant, the aPRs were consistently more protective for joint than individual exposures across helminths.

I plotted the cluster-level prevalence of each helminth across the observed range of cluster-level deworming coverage (Figure 2.3), cluster-level hygienic latrine coverage (Figure A.2), and cluster-level finished floor coverage (Figure A.3). For *Ascaris* and hookworm, the prevalence stayed approximately the same across different levels of cluster deworming coverage. The prevalence of *Trichuris* increased as cluster deworming and sanitation coverage increased. *Ascaris* prevalence was nearly constant across the range of cluster-level sanitation coverage. Hookworm prevalence decreased as sanitation coverage increased. For *Trichuris* there was a slight increase in the cluster-level prevalence around 40% cluster-level deworming coverage, but overall no substantial increase. The cluster-level prevalence of *Ascaris* decreased from around 15% to 8% as cluster-level finished floor coverage increased from 0% to 60%. Overall there appeared to be no association between cluster-level prevalence of hookworm and *Trichuris* and cluster-level finished floor coverage. The village cluster level intraclass correlation coefficients were 0.11 (95% CI 0.08, 0.16) for *Ascaris*, 0.02 (95% CI 0.00, 0.05) for hookworm, and 0.21 (95% CI 0.15, 0.27) for *Trichuris*.

Table 2.3: Adjusted prevalence ratios and measures of interaction between hygienic latrines and deworming

		Ascaris		Hookworm		*Trichuris*	
Deworming	Hygienic Latrine	n/N	aPR* (95% CI)	n/N	aPR* (95% CI)	n/N	aPR* (95% CI)
-	-	106/652	1.00 (–,–)	47/652	1.00 (–,–)	130/652	1.00 (–,–)
+	-	49/445	0.63 (0.46,0.87)	23/445	0.82 (0.50,1.32)	107/445	1.20 (0.95,1.51)
-	+	44/305	0.97 (0.69,1.37)	11/305	0.55 (0.28,1.07)	70/305	1.20 (0.91,1.58)
+	+	14/219	0.44 (0.25,0.77)	9/219	0.81 (0.38,1.73)	35/219	0.87 (0.61,1.24)
	Measures of interaction						
	RERI†		-0.43 (-2.40,0.48)		0.45 (-0.88,1.15)		-0.65 (-1.41,-0.13)
	RPR‡		0.77 (0.35,1.53)		1.53 (0.54,4.33)		0.59 (0.35,0.92)

*PRs adjusted for age, sex, sub-district, household wealth, cluster-level wealth, and mother's education level

†Relative excess risk due to interaction (RERI). A RERI=0 indicates no interaction on the additive scale, RERI>0 indicates synergistic interaction on the additive scale for monotonic aPRs, RERI>1 indicating synergistic interaction on the additive scale for non-monotonic aPRs, and RERI<0 indicates antagonistic interaction on the additive scale.

‡Ratio of prevalence ratios

Table 2.4: Adjusted prevalence ratios and measures of interaction between finished floors and deworming

		Ascaris		Hookworm		*Trichuris*	
Deworming	Finished floor	n/N	aPR* (95% CI)	n/N	aPR* (95% CI)	n/N	aPR* (95% CI)
-	-	141/847	1.00 (–,–)	55/847	1.00 (–,–)	177/847	1.00 (–,–)
+	-	59/572	0.60 (0.45,0.80)	31/572	0.94 (0.61,1.44)	126/572	1.05 (0.85,1.29)
-	+	9/110	0.54 (0.28,1.03)	3/110	0.50 (0.15,1.70)	23/110	1.10 (0.72,1.68)
+	+	4/93	0.29 (0.11,0.79)	1/93	0.26 (0.03,1.98)	16/93	0.96 (0.59,1.54)
	Measures of interaction						
	RERI†		0.51 (-3.65,2.00)		-1.04 (–,–)		-0.23 (-1.50,0.37)
	RPR‡		1.13 (0.25,3.67)		0.91 (–,–)		0.83 (0.41,1.55)

*PRs adjusted for age, sex, sub-district, household wealth, cluster-level wealth, and mother's education level

†Relative excess risk due to interaction (RERI). A RERI=0 indicates no interaction on the additive scale, RERI>0 indicates synergistic interaction on the additive scale for monotonic aPRs, RERI>1 indicating synergistic interaction on the additive scale for non-monotonic aPRs, and RERI<0 indicates antagonistic interaction on the additive scale.

‡Ratio of prevalence ratios

Figure 2.2: STH prevalence by exposure to deworming, hygienic latrines, and finished floors

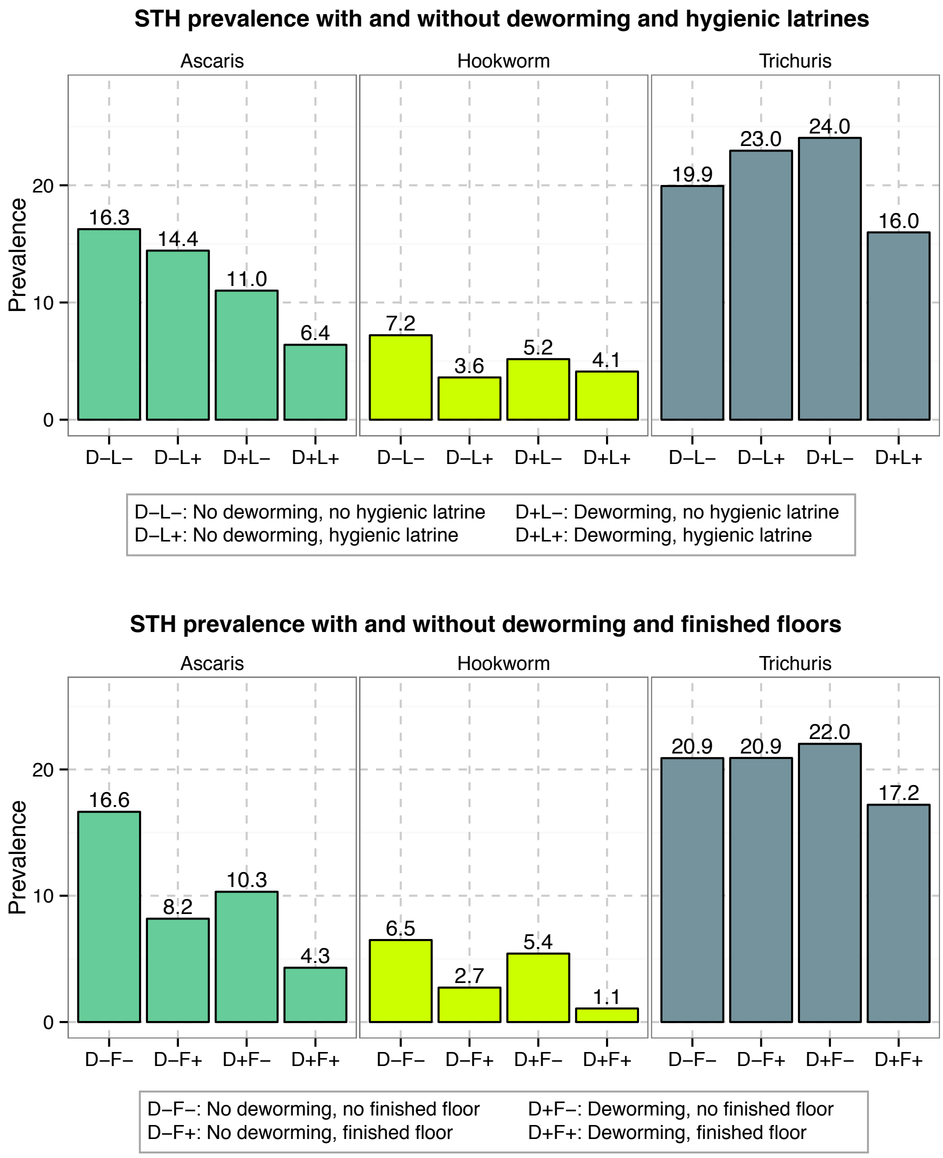
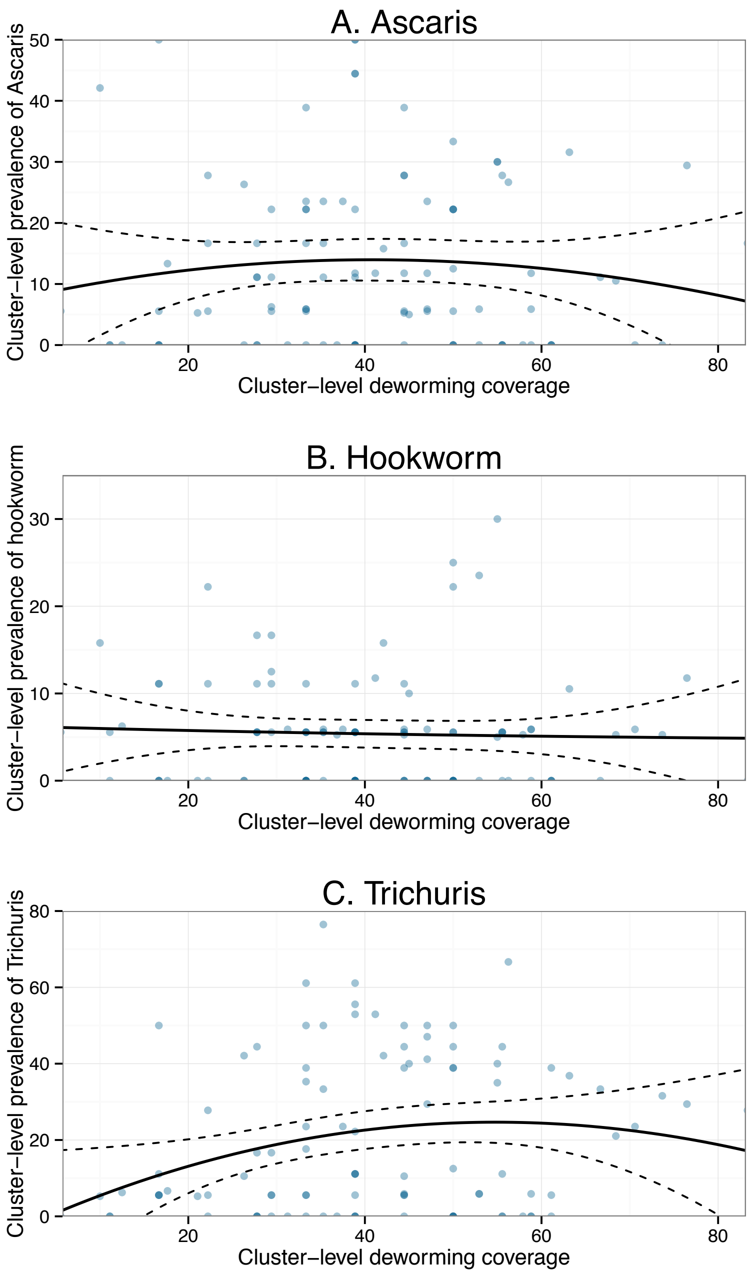


Figure 2.3: Cluster-level STH prevalence by cluster-level deworming coverage



Observed data points for cluster-level prevalence and cluster-level deworming coverage (blue points), LOWESS fit (solid line), and point-wise 95% confidence bands (dashed lines).

2.4 Discussion

In this cross-sectional study of rural, low-income households in Bangladesh, where national school-based deworming has been implemented for the last five years, prevalences among school-aged children were 14% for *Ascaris*, 8% for hookworm, and 21% for *Trichuris*; 40% of school-aged children had at least one STH infection, and 12% had multiple infections. In addition, 26% of children 1-4 years and 30% of women of childbearing age were infected with at least one STH. There were very few moderate/heavy STH infections in the study population. Approximately half of children under 15 years were reported to have been dewormed in the prior six months. When I estimated associations between individual exposures (deworming, hygienic latrines, and finished floors) and STH prevalence, I found protective adjusted prevalence ratios for *Ascaris* and hookworm but nearly null associations for *Trichuris*. When I considered joint exposure to deworming with hygienic latrines or finished floors, I found that across helminths the joint exposures were consistently associated with a lower prevalence than individual exposures. Contrary to what I would expect based on transmission models, I found that STH prevalence was approximately the same across levels of cluster-level deworming coverage or sanitation coverage, suggesting that household level access to deworming or improved environmental conditions was more important than cluster-level coverage in this setting. These findings support further exploration of improved sanitation and finished floors as interventions to complement MDA to sustainably reduce STH transmission in Bangladesh.

Because this study was cross-sectional, I have estimated associations between exposures and STH prevalence; because I did not measure incident infections, I could not assess the effect of exposures on incidence and transmission of STH. A limitation of this design is that the exposure measurement occurred following outcome measurement. This is suboptimal because it is possible that the outcome status of an individual would trigger a change in their exposure, leading to reverse causation. However, for the exposures of interest – access to a hygienic latrine and finished floors – I consider it highly unlikely that the respondents' exposure status changed between October and December 2012. Furthermore, respondents were unaware of their outcome status throughout the study so a response to the October measurement seems highly improbable. While the study population included both SHEWA-B intervention and control areas without the SHEWA-B intervention, there was a range of exposures across both groups and do not consider intervention group status to be a potential confounder. Furthermore, the interim evaluation of SHEWA-B found that the intervention did not strongly improve health behaviors or child health [18]. I assumed that the sanitation and flooring infrastructure available at the time of stool sample collection were present when deworming occurred six months earlier. I consider this to be a reasonable assumption, however, if this assumption was not true sanitation and flooring exposures may have been misclassified, prevalence ratios would be biased towards the null and the effects on measures of interaction would be unpredictable [51].

To estimate cluster-level coverage, I averaged across approximately 18 households per cluster, and most households were within a few minutes walking distance from each other. I averaged across the three age groups sampled, and typically there were six people per age group per cluster. A larger sample of individuals per age group in each cluster would have likely provided more accurate and precise estimates of cluster-level sanitation and deworming conditions.

We collected stool samples once per individual; in the context of MDA, it has been shown that a single stool sample is sometimes insufficient to detect STH infection using Kato-Katz and that ideally multiple serial samples should be collected per individual [52]; as a result, prevalence may be underestimated in this study. However, Knopp et al. found that FLOTAC on single stool sample was more sensitive than Kato-Katz on three serial stool samples [53]. While such validation has not yet been done for mini-FLOTAC, given the similarity of mini-FLOTAC and FLOTAC, it is likely that findings may be similar, and sensitivity may have been high in this study even with a single stool sample.

Another limitation of this study is that one of the main exposures of interest, deworming in the past six months, was self-reported. In a study of recall of disease symptoms and medication use in Kenya, Feikin et al. found that recall of anti-malarial and antibiotic use decreased as the number of days since visiting a health clinic increased [54]. They argue that recall of medication will be under-reported with longer recall periods. However, it is also possible that reporting was subject to courtesy bias so that deworming was over-reported. Given that it is unlikely that respondents knew whether or not they or their children had an STH infection at the time of the interview, I posit that any misclassification of deworming use likely did not differ by STH infection status; thus, if non-differential misclassification occurred, the point estimates would be biased towards the null [55]. However, given the relatively long recall period for deworming, it is possible that individuals who were dewormed in the prior six months were reinfected prior to stool collection in this study. For this reason and because of the possible misclassification of deworming due to poor recall, the associations between deworming and STH infection do not necessarily measure the reduction in STH attributable directly to deworming.

In this study population, moderate and heavy intensity infections were rare; since intensity of infection drives the rate of transmission, this finding suggests that transmission may be waning in this study population and that the MDA program in Bangladesh may have introduced a new steady state of transmission [29]. I also found that prevalence of any STH infection among school-aged children was 40% compared to the prevalence of 80% reported by the Ministry of Health and Family Welfare prior to the initiation of school-based deworming [2]. The prevalence observed is consistent with studies of school-based deworming with one-year follow-up and very high coverage [56–58]. Prevalence is typically expected to be lower among young children and adults, and transmission theory [20–22] and empirical findings [30, 31] suggest that prevalence decreases in pre-school age and adult populations when coverage is high. I found that 26% of children 1-4 years and 30% of women of childbearing age had an STH infection. The similar prevalence in these two groups to prevalence among school-aged children may suggest that school-based deworming alone might be insufficient to interrupt transmission.

When I explored associations with STH infection and deworming, finished floors, and hygienic latrines, I found protective associations of each exposure with *Ascaris* and hookworm but no associations between the individual exposures and *Trichuris*. The association with deworming was statistically significant and protective for *Ascaris* but was not statistically significant for hookworm or *Trichuris*. These findings are consistent with those of randomized controlled trials, which have found cure rates for single dose mebendazole and albendazole are greater than 90% for *Ascaris*, 30-90% for hookworm, and around 40% for *Trichuris* [59–61]. I found protective but not statistically significant adjusted associations with hygienic latrines for *Ascaris* and hookworm and no association with *Trichuris*. In a meta-analysis, Ziegelbauer et al. estimated that the odds ratio for availability of toilets of any kind was 0.46 (95% CI 0.33, 0.64) for *Ascaris*, 0.56 (95% CI 0.46, 0.70) for *Trichuris*, and 0.58 (95% CI 0.45, 0.76) for hookworm [6]. These results may differ due to differences in exposure definition. I observed a strong protective association with living in a household with a finished floor for *Ascaris* and hookworm, although the result was not statistically significant for hookworm. There was no association between finished floors and *Trichuris*. The findings related to finished floors are consistent with findings in the literature [8–10], although few studies have examined the association with all three helminths considered here. The null associations for all three individual exposures with *Trichuris*, the most prevalent helminth in this population, are noteworthy. There is some evidence that the prevalence of *Trichuris* decreases more slowly than that of other STH and that reinfection with *Trichuris* occurs more rapidly following intervention [62–64]. This may be because of longer survival of adult worms or because *Trichuris* has a higher reproductive rate than other STH [62, 63]. The higher observed prevalence of *Trichuris* compared to *Ascaris* and hookworm in this study likely reflects these parasite-specific differences in biology and response to intervention.

I assessed the potential for interaction among deworming, hygienic latrines, and finished floors, and I found a consistent pattern across organisms that suggests possible synergistic interactions. Across helminths and combinations of exposures, aPRs for joint exposures were consistently more protective than those for individual exposures. The sample size was powered to estimate prevalence but not to estimate interactions between exposures. Thus, my estimates of the RERI and RPR were in most cases underpowered – particularly for improved floors, which were relatively rare in this population. Evidence of interaction on the additive scale can suggest causal interaction [47]; however, due to the cross-sectional nature of the design, I cannot attribute this finding to causation. Nevertheless, the results support further exploration of these interactions using a prospective design that is powered to explore interaction. One would expect based on transmission theory that increasing coverage of deworming, improved sanitation, and finished floors at the cluster level would be associated with decreased STH prevalence [20–22]. For example, MDA typically targets school-aged children and not the whole community because school-aged children usually have a higher burden of infection and thus drive transmission in communities [21]. Due to herd effects, I would expect MDA to reduce the burden of STH not only among school-aged children but also among younger and older age groups. Following deworming, herd effects occur when the decreased shedding of infective stages in feces into the environment reduces transmission community-wide [22]. Such herd effects are also biologically plausible for improved sanitation and finished floors. Contrary to what has been predicted by transmission models [20–22], I found that individual deworming and household sanitation had a stronger protective association than living in a village cluster with high deworming or sanitation coverage. Yet, I also found relatively large intraclass correlation coefficients (ICCs) at the village level for *Trichuris* (ICC=0.25) and a moderate ICC for *Ascaris* (ICC=0.11); these findings are similar to ICCs reported in the literature [65]. For *Trichuris*, the most prevalent STH in this population, the ICC indicates that approximately a quarter of the variance can be explained by village membership. Thus, village membership appears to be an important predictor of STH infection, but cluster-level deworming, hygienic latrine, and finished flooring coverage do not appear to be the most important predictors at the cluster level. It is possible that my finding is specific to the prevalence level in this population; if prevalence were higher, herd effects might be stronger, and community-level exposures might have a stronger association with community prevalence.

A large body of evidence describes clustering of STH infection within the household [10, 23–26, 28, 66]; such clustering could reflect shared environmental exposures, hygiene behaviors, infrastructure (e.g. flooring), or genetics and immunological responses. Moraes and Cairncross found that the extent of household clustering of STH infection depended upon the household's access to drainage and sewerage and that household clustering was stronger in communities where community-level drainage was in place than in communities without drainage [27]. Their finding may suggest that when community level intervention coverage is high, household transmission dominates, whereas when community level coverage is low, both forms of transmission are important. Such a pattern may explain the lack of association I found between village-level coverage of interventions and village-level prevalence. Alternatively, it is possible that community-level exposures are stronger drivers of prevalence and transmission in a high prevalence setting. This could be because herd effects may be stronger when community-level prevalence is higher, yet when it is lower, household level exposures may be more strongly associated with infection than exposures in the greater community.

One potential factor I did not explore that could affect the extent to which cluster-level coverage is associated with STH prevalence is population density. In dense villages where households are located very close together, the probability of transmission resulting from other people's unimproved sanitation rather than one's own sanitation is higher [67]. Similarly, the impact of high village-level deworming coverage may be stronger in densely populated villages than in low-density villages since the extent of transmission between individuals is likely to be greater in higher density settings. Thus, population density may be an important effect modifier of the relationship between sanitation, deworming, and STH infection. Indeed, in an analysis considering population density and urban extents, Pullan et al. found that hookworm prevalence was greatly reduced in urban areas and that *Ascaris* and *Trichuris* prevalence were higher in slums in peri-urban areas [68]. Halpenny et al. found that in rural Panama chronic *Giardia* spp., *Entamoeba histolytica*, and *Entamoeba dispar* infections were associated with living in a higher density area, but spatial clustering of STH infections was associated with lower household density [69, 70]. Further work is needed to explore the role of population density and the extent to which it modifies the effect of deworming and sanitation exposures or interventions on STH infection.

In summary, I found that STH infections were prevalent among school-aged children (despite targeting by school-based MDA for the last five years) and also among pre-school aged children and women of childbearing age in the same community. There was evidence that individuals who were dewormed and had access to hygienic latrines and finished flooring in their household had a lower prevalence of STH than those with deworming alone. These results suggest that coupling MDA with sanitation and flooring interventions to yield greater or more sustained reductions in prevalence of STH infection is a strategy that should be evaluated rigorously, perhaps in a randomized trial. By randomizing the provision of deworming, sanitation, and flooring interventions and measuring infection prospectively, such a trial would be able to attribute reductions in STH incidence and intensity to particular interventions and could compare reinfection rates following deworming under sanitation versus flooring interventions. Considering the growing concerns about the potential for resistance to anthelminthics [71, 72], provision of sanitation and flooring are promising complementary interventions with the potential to more sustainably reduce STH transmission. The research and programming community focusing on neglected tropical diseases including STH have largely focused on preventive chemotherapy, and the water sanitation and hygiene sector has largely worked independently of the neglected tropical disease sector [11]. This study is one of the first to examine the independent and combined effects of exposures from each sector, and the findings suggest the need for further intersectoral collaboration and exploration of sanitation and flooring as complementary interventions to deworming.

Bibliography

[1] World Health Organization. *Deworming for health and development: report of the Third Global Meeting of the Partners for Parasite Control.* 2005.

[2] Ministry of Health and Family Welfare. *A Situation Analysis: Neglected Tropical Diseases in Bangladesh.* 2010. URL: `http://pdf.usaid.gov/pdf_docs/pnady849.pdf`.

[3] Jia TW, Melville S, Utzinger J, King CH, and Zhou XN. Soil-transmitted helminth reinfection after drug treatment: a systematic review and meta-analysis. *PLoS Neglected Tropical Diseases* 6:5 (2012).

[4] Barreto ML, Genser B, Strina A, Teixeira MG, Assis AMO, Rego RF, Teles CA, Prado MS, Matos SM, Alcântara-Neves NM, et al. Impact of a citywide sanitation program in Northeast Brazil on intestinal parasites infection in young children. *Environmental Health Perspectives* 118:11 (2010), p. 1637.

[5] Mascarini-Serra LM, Telles CA, Prado MS, Mattos SA, Strina A, Alcantara-Neves NM, and Barreto ML. Reductions in the prevalence and incidence of geohelminth infections following a city-wide sanitation program in a Brazilian Urban Centre. *PLoS Neglected Tropical Diseases* 4:2 (2010), e588.

[6] Ziegelbauer K, Speich B, Mäusezahl D, Bos R, Keiser J, and Utzinger J. Effect of sanitation on soil-transmitted helminth infection: systematic review and meta-analysis. *PLoS Medicine* 9:1 (2012), e1001162.

[7] Freeman MC, Clasen T, Brooker SJ, Akoko DO, and Rheingans R. The Impact of a School-Based Hygiene, Water Quality and Sanitation Intervention on Soil-Transmitted Helminth Reinfection: A Cluster-Randomized Trial. *The American Journal of Tropical Medicine and Hygiene* 89:5 (2013), pp. 875–883.

[8] Faust EC and Giraldo LE. Parasitological surveys in Cali, Departamento del Valle, Colombia VI. Strongyloidiasis in Barrio Siloé, Cali, Colombia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 54:6 (1960), pp. 556–563.

[9] Hall A, Conway DJ, Anwar KS, and Rahman ML. Strongyloides stercoralis in an urban slum community in Bangladesh: factors independently associated with infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 88:5 (1994), pp. 527–530.

[10] Walker M, Hall A, and Basáñez MG. Individual predisposition, household clustering and risk factors for human infection with Ascaris lumbricoides: new epidemiological insights. *PLoS Neglected Tropical Diseases* 5:4 (2011), e1047.

[11] Freeman MC, Ogden S, Jacobson J, Abbott D, Addiss DG, Amnie AG, Beckwith C, Cairncross S, Callejas R, Colford Jr JM, et al. Integration of water, sanitation, and hygiene for the prevention and control of neglected tropical diseases: a rationale for inter-sectoral collaboration. *PLoS Neglected Tropical Diseases* 7:9 (2013), e2439.

[12] Freeman MC, Greene LE, Dreibelbis R, Saboori S, Muga R, Brumback B, and Rheingans R. Assessing the impact of a school-based water treatment, hygiene and sanitation programme on pupil absence in Nyanza Province, Kenya: a cluster-randomized trial. *Tropical Medicine & International Health* 17:3 (2012), pp. 380–391.

[13] Hesham Al-Mekhlafi M, Surin J, Atiya A, Ariffin W, Mohammed Mahdy A, and Che Abdullah H. Pattern and predictors of soil-transmitted helminth reinfection among aboriginal schoolchildren in rural Peninsular Malaysia. *Acta Tropica* 107:2 (2008), pp. 200–204.

[14] Henry FJ. Reinfection with Ascaris lumbricoides after chemotherapy: a comparative study in three villages with varying sanitation. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 82:3 (1988), pp. 460–464.

[15] Cort W, Schapiro L, and Stoll N. A study of reinfection after treatment with hookworm and Ascaris in two villages in Panama. *American Journal of Epidemiology* 10:3 (1929), pp. 614–625.

[16] Cattaneo MD, Galiani S, Gertler PJ, Martinez S, and Titiunik R. Housing, health, and happiness. *American Economic Journal: Economic Policy* (2009), pp. 75–105.

[17] Blot WJ and Day NE. Synergism and interaction: are they equivalent? *American Journal of Epidemiology* 110:1 (1979), pp. 99–100.

[18] Huda TMN, Unicomb L, Johnston RB, Halder AK, Yushuf Sharker MA, and Luby SP. Interim evaluation of a large scale sanitation, hygiene and water improvement programme on childhood diarrhea and respiratory disease in rural Bangladesh. *Social Science & Medicine* 75:4 (2012), pp. 604–611.

[19] Hadi A. A participatory approach to sanitation: experience of Bangladeshi NGOs. *Health Policy and Planning* 15:3 (2000), pp. 332–337.

[20] Anderson RM and May RM. Helminth infections of humans: mathematical models, population dynamics, and control. *Advances in Parasitology* 24 (1985), pp. 1–101.

[21] Anderson R and Medley G. Community control of helminth infections of man by mass and selective chemotherapy. *Parasitology* 90:04 (1985), pp. 629–660.

[22] Anderson RM, Truscott JE, Pullan RL, Brooker SJ, and Hollingsworth TD. How effective is school-based deworming for the community-wide control of soil-transmitted helminths? *PLoS Neglected Tropical Diseases* 7:2 (2013), e2027.

[23] Brooker S, Alexander N, Geiger S, Moyeed RA, Stander J, Fleming F, Hotez PJ, Correa-Oliveira R, and Bethony J. Contrasting patterns in the small-scale heterogeneity of human helminth infections in urban and rural environments in Brazil. *International Journal for Parasitology* 36:10 (2006), pp. 1143–1151.

[24] Forrester J, Scott M, Bundy D, and Golden M. Clustering of Ascaris lumbricoides and Trichuris trichiura infections within households. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 82:2 (1988), pp. 282–288.

[25] Conway DJ, Hall A, Anwar KS, Rahman ML, and Bundy DA. Household aggregation of Strongyloides stercoralis infection in Bangladesh. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 89:3 (1995), pp. 258–261.

[26] Lindo JF, Robinson RD, Terry SI, Vogel P, Gam AA, Neva FA, and Bundy DA. Age-prevalence and household clustering of Strongyloides stercoralis infection in Jamaica. *Parasitology* 110:01 (1995), pp. 97–102.

[27] Moraes LRS and Cairncross S. Environmental interventions and the pattern of geohelminth infections in Salvador, Brazil. *Parasitology* 129:02 (2004), pp. 223–232.

[28] Pullan RL, Bethony JM, Geiger SM, Cundill B, Correa-Oliveira R, Quinnell RJ, and Brooker S. Human helminth co-infection: analysis of spatial patterns and risk factors in a Brazilian community. *PLoS Neglected Tropical Diseases* 2:12 (2008), e352.

[29] Bundy D and Medley G. Immuno-epidemiology of human geohelminthiasis: ecological and immunological determinants of worm burden. *Parasitology* 104:S1 (1992), S105–S119.

[30] Bundy DA, Wong MS, Lewis LL, and Horton J. Control of geohelminths by delivery of targeted chemotherapy through schools. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 84:1 (1990), pp. 115–120.

[31] Asaolu S, Holland C, and Crompton D. Community control of Ascaris lumbricoides in rural Oyo State, Nigeria: mass, targeted and selective treatment with levamisole. *Parasitology* 103:02 (1991), pp. 291–298.

[32] Miguel E and Kremer M. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72:1 (2004), pp. 159–217.

[33] Luby SP, Halder AK, Huda T, Unicomb L, and Johnston RB. The effect of handwashing at recommended times with water alone and with soap on child diarrhea in rural Bangladesh: an observational study. *PLoS Medicine* 8:6 (2011), e1001052.

[34] Barda BD, Rinaldi L, Ianniello D, Zepherine H, Salvo F, Sadutshang T, Cringoli G, Clementi M, and Albonico M. Mini-FLOTAC, an innovative direct diagnostic technique for intestinal parasitic infections: experience from the field. *PLoS Neglected Tropical Diseases* 7:8 (2013), e2344.

[35] Barda B, Zepherine H, Rinaldi L, Cringoli G, Burioni R, Clementi M, and Albonico M. Mini-FLOTAC and Kato-Katz: helminth eggs watching on the shore of lake Victoria. *Parasites & Vectors* 6:1 (2013), p. 220.

[36] WHO Expert Committee. Prevention and control of schistosomiasis and soil-transmitted helminthiasis. *World Health Organization Technical Report Series* 912 (2002).

[37] World Health Organization, UNICEF, et al. Core questions on drinking water and sanitation for household surveys (2006). URL: `http://www.who.int/water_sanitation_health/monitoring/household_surveys/en/`.

[38] Greenland S, Pearl J, and Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 10:1 (1999), pp. 37–48.

[39] Vyas S and Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. *Health Policy and Planning* 21:6 (2006), pp. 459–468.

[40] Patil SR, Arnold BF, Salvatore A, Briceno B, Colford Jr JM, and Gertler PJ. A randomized, controlled study of a rural sanitation behavior change program in Madhya Pradesh, India. *World Bank Policy Research Working Paper* 6702 (2013).

[41] Hastie T, Tibshirani R, Friedman J, and Franklin J. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer* 27:2 (2005), pp. 83–85.

[42] Zou G. A modified poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* 159:7 (2004), pp. 702–706.

[43] van der Laan MJ, Polley EC, et al. Super learner. *Statistical Applications in Genetics and Molecular Biology* 6:1 (2007), pp. 1–23.

[44] Rothman KJ, Greenland S, and Walker AM. Concepts of interaction. *American Journal of Epidemiology* 112:4 (1980), pp. 467–470.

[45] Saracci R. Interaction and synergism. *American Journal of Epidemiology* 112:4 (1980), pp. 465–466.

[46] Rothman KJ. *Modern Epidemiology.* Boston: Little, Brown, 1986.

[47] Rothman KJ, Greenland S, and Lash TL. *Modern epidemiology.* Lippincott Williams & Wilkins, 2008.

[48] Knol MJ and VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. *International Journal of Epidemiology* 41:2 (2012), pp. 514–520.

[49] VanderWeele TJ and Robins JM. The identification of synergism in the sufficient-component-cause framework. *Epidemiology* 18:3 (2007), pp. 329–339.

[50] VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiology* 20:1 (2009), pp. 6–13.

[51] Greenland S. The effect of misclassification in the presence of covariates. *American Journal of Epidemiology* 112:4 (1980), pp. 564–569.

[52] Knopp S, Mgeni AF, Khamis IS, Steinmann P, Stothard JR, Rollinson D, Marti H, and Utzinger J. Diagnosis of soil-transmitted helminths in the era of preventive chemotherapy: effect of multiple stool sampling and use of different diagnostic techniques. *PLoS Neglected Tropical Diseases* 2:11 (2008), e331.

[53] Knopp S, Rinaldi L, Khamis IS, Stothard JR, Rollinson D, Maurelli MP, Steinmann P, Marti H, Cringoli G, and Utzinger J. A single FLOTAC is more sensitive than triplicate Kato–Katz for the diagnosis of low-intensity soil-transmitted helminth infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 103:4 (2009), pp. 347–354.

[54] Feikin DR, Audi A, Olack B, Bigogo GM, Polyak C, Burke H, Williamson J, and Breiman RF. Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. *International Journal of Epidemiology* 39:2 (2010), pp. 450–458.

[55] Hutcheon JA, Chiolero A, and Hanley JA. Random measurement error and regression dilution bias. *BMJ: British Medical Journal (Overseas & Retired Doctors Edition)* 340:7761 (2010).

[56] Albonico M, Shamlaye N, Shamlaye C, and Savioli L. Control of intestinal parasitic infections in Seychelles: a comprehensive and sustainable approach. *Bulletin of the World Health Organization* 74:6 (1996), p. 577.

[57] Albonico M, Stoltzfus RJ, Savioli L, Chwaya HM, d'Harcourt E, and Tielsch JM. A controlled evaluation of two school-based anthelminthic chemotherapy regimens on intensity of intestinal helminth infections. *International Journal of Epidemiology* 28:3 (1999), pp. 591–596.

[58] Massa K, Magnussen P, Sheshe A, Ntakamulenga R, Ndawi B, and Olsen A. The effect of the community-directed treatment approach versus the school-based treatment approach on the prevalence and intensity of schistosomiasis and soil-transmitted helminthiasis among schoolchildren in Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 103:1 (2009), pp. 31–37.

[59] Albonico M, Smith PG, Hall A, Chwaya HM, Alawi KS, and Savioli L. A randomized controlled trial comparing mebendazole and albendazole against Ascaris, Trichuris and hookworm infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 88:5 (1994), pp. 585–589.

[60] Vercruysse J, Behnke JM, Albonico M, Ame SM, Angebault C, Bethony JM, Engels D, Guillard B, Hoa NTV, Kang G, et al. Assessment of the anthelmintic efficacy of albendazole in school children in seven countries where soil-transmitted helminths are endemic. *PLoS Neglected Tropical Diseases* 5:3 (2011), e948.

[61] Steinmann P, Utzinger J, Du ZW, Jiang JY, Chen JX, Hattendorf J, Zhou H, and Zhou XN. Efficacy of single-dose and triple-dose albendazole and mebendazole against soil-transmitted helminths and Taenia spp.: a randomized controlled trial. *PLoS One* 6:9 (2011), e25003.

[62] Bundy DA. Epidemiological aspects of Trichuris and trichuriasis in Caribbean communities. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 80:5 (1986), pp. 706–718.

[63] Bundy DA, Thompson DE, Cooper ES, Golden MH, and Anderson RM. Population dynamics and chemotherapeutic control of Trichuris trichiura infection of children in Jamaica and St. Lucia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 79:6 (1985), pp. 759–764.

[64] Anderson RM. The population dynamics and epidemiology of intestinal nematode infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 80:5 (1986), pp. 686–696.

[65] Moncayo AL, Vaca M, Amorim L, Rodriguez A, Erazo S, Oviedo G, Quinzo I, Padilla M, Chico M, Lovato R, et al. Impact of long-term treatment with ivermectin on the prevalence and intensity of soil-transmitted helminth infections. *PLoS Neglected Tropical Diseases* 2:9 (2008), e293.

[66] Otto G, Cort W, and Keller A. Environmental studies of families in Tennessee infested with Ascaris, Trichuris and hookworm. *American Journal of Epidemiology* 14:1 (1931), pp. 156–193.

[67] Brooker S, Clements AC, and Bundy DA. Global epidemiology, ecology and control of soil-transmitted helminth infections. *Advances in Parasitology* 62 (2006), pp. 221–261.

[68] Pullan RL and Brooker SJ. The global limits and population at risk of soil-transmitted helminth infections in 2010. *Parasites & Vectors* 5:1 (2012), p. 81.

[69] Halpenny CM, Koski KG, Valdés VE, and Scott ME. Prediction of child health by household density and asset-based indices in impoverished Indigenous villages in rural Panamá. *The American Journal of Tropical Medicine and Hygiene* 86:2 (2012), p. 280.

[70] Halpenny CM, Paller C, Koski KG, Valdés VE, and Scott ME. Regional, household and individual factors that influence soil transmitted helminth reinfection dynamics in preschool children from rural indigenous Panama. *PLoS Neglected Tropical Diseases* 7:2 (2013), e2070.

[71] Albonico M, Engels D, and Savioli L. Monitoring drug efficacy and early detection of drug resistance in human soil-transmitted nematodes: a pressing public health agenda for helminth control. *International Journal for Parasitology* 34:11 (2004), pp. 1205–1210.

[72] Geerts S and Gryseels B. Anthelmintic resistance in human helminths: a review. *Tropical Medicine & International Health* 6:11 (2001), pp. 915–921.

Chapter 3

Assessment of a national-scale water, sanitation and hygiene intervention in rural Bangladesh: Measuring the effect of implementation quality

3.1 Background

Scientists and development stakeholders argue that health interventions proven effective in randomized efficacy trials should be translated into large-scale programs to benefit public health [1–4]. Substantive evidence supports the scale-up of numerous health interventions [5–7], and since the establishment of the Millennium Development Goals (MDGs) the funding and motivation for such scale-up has grown [2, 8–10]. Translating interventions shown to be efficacious on a small scale to large-scale programs presents implementation challenges, and systematically studying the reasons for scale-up success or failure is essential to refine and sustain public health programs [11–14]. A growing body of literature documents barriers to and facilitators of scale-up, yet there currently is little empirical evidence about how best to scale up [9, 13, 15–22]. A systematic review of such models advocated a data-based approach to determining constraints to and facilitators of scale-up [15].

In low-income countries, enteric infections continue to account for one of the largest disease burdens among young children, and dozens of efficacy studies have demonstrated that in trial conditions decentralized water, sanitation, and hygiene (WASH) interventions can dramatically reduce enteric infection risk [23–34]. The few existing rigorous evaluations of large-scale WASH interventions found no effect on access to improved sanitation and mixed results related to handwashing and diarrhea prevalence [35–37]. Empirical evidence about the implementation of large-scale WASH interventions would improve our understanding of how best to scale-up and contribute to reduced diarrheal disease and mortality and may contribute to improved scale-up of interventions in other sectors. As governments and stakeholders begin to deliver WASH and other intervention programs at a national scale, there is a scientific imperative to evaluate program impacts on health and to document reasons for intervention success or failure [11, 38].

One of the largest WASH interventions in a low-income country to date is the Sanitation Hygiene Education and Water Supply in Bangladesh (SHEWA-B) program, which was implemented by UNICEF and the Government of Bangladesh. SHEWA-B targeted approximately 20.4 million beneficiaries from 2007 to 2012. The intervention aimed to promote hygiene practices and reduce diarrhea and other water and hygiene-related diseases among the poorest in rural Bangladesh. UNICEF and the Government of Bangladesh partnered with local government institutes and a large network of local non-governmental organizations (NGOs), which recruited local residents to serve as community hygiene promoter (CHPs) and provided them with training and supervision.

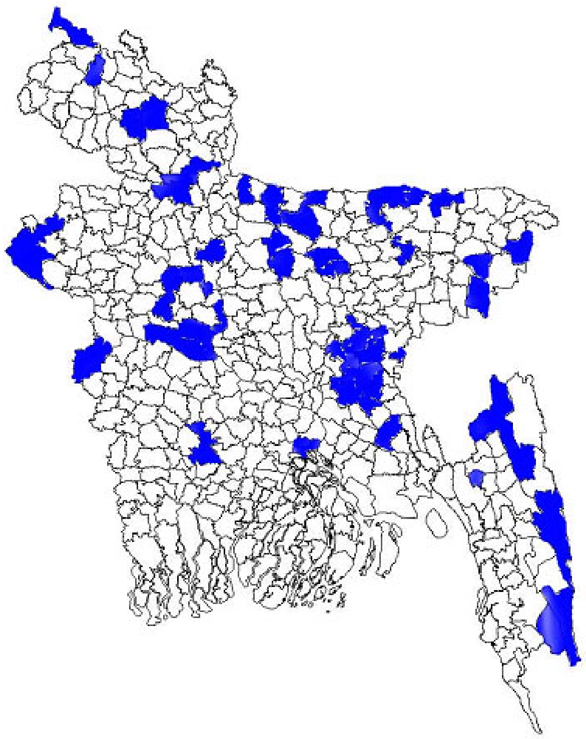
An interim assessment of SHEWA-B in 2009 found little to no improvement in measures of behavior or child health, such as handwashing, or the prevalence of diarrhea and respiratory illness among children under five years [39]. These results could reflect a suboptimal intervention that needed to be better tailored to the target population or an appropriate intervention that needed to be better implemented. This study's objective was to estimate the extent to which hygiene behavior and conditions might have improved if SHEWA-B had been better implemented. Such information can be used by UNICEF and other similar organizations to improve the design and/or implementation of future large-scale WASH interventions. I expect that the assessment methods in this chapter will have broader application for the assessment of large-scale program implementation quality beyond the WASH sector.

3.2 Methods

The SHEWA-B Intervention

UNICEF implemented SHEWA-B in under-served areas without other large-scale WASH interventions in 2007 (Figure 3.1). UNICEF and the Bangladesh Department of Public Health Engineering (DPHE) recruited local NGOs that recruited CHPs from the communities that they worked in. CHPs received 15 days of initial training in 2007 and 12 total days of refresher between 2009 and 2012. UNICEF partnered with international and Bangladesh NGOs including Water Aid Bangladesh, Plan International, and Dhaka Ahsania Mission to train, coach and monitor local NGOs. At the time of this study, each CHP was responsible to conduct household visits and community meetings for approximately 1200-1500 households every three months. There was little migration in the SHEWA-B target population over the course of the intervention, and typically target beneficiaries (mothers of children under five) were available during the day to participate in CHP-run activities. CHPs received an incentive of 140 Bangladeshi Taka approximately 1.80 US dollar per day, which is roughly half the daily wage of an unskilled laborer. In response to the interim assessment results in 2009, UNICEF attempted to address major problems in delayed CHP funds disbursement, which may have reduced field implementation activities (we provide additional details on UNICEF's response to the interim assessment in the appendix). SHEWA-B installed around 17,606 new safe water points among under-served and un-served rural communities. Non-governmental organizations were responsible for motivating the intervention communities to develop plans to install latrines. Further details about the SHEWA-B intervention are provided elsewhere [39].

Figure 3.1: Map of SHEWA-B implementation areas



Blue areas indicate upazilas (sub-districts) in which SHEWA-B was implemented, and white areas indicate upazilas in which it was not implemented.

Cross-sectional survey

A cross-sectional study was implemented in a sample of intervention villages between June 2011 and April 2012 by 142 field staff trained by ICDDR,B. Individuals were eligible to participate in the survey if at least one child under five years resided in their household. Data were collected within village-level clusters. The questionnaire measured UNICEF target outcomes, such as respondents' self-reported hygiene and sanitation practices. Field staff conducted spot checks and asked respondents about their health behaviors. Respondents were asked if any CHPs visited their homes, and if so, the name of the CHP, how often s/he visited, and what topics they discussed. Staff attempted to interview each CHP in the sampled clusters about how often they visited households and conducted community meetings and asked them to recall key SHEWA-B messages.

Focus groups and in-depth interviews

Field staff collected qualitative data to identify barriers to SHEWA-B promoted practices and CHPs' work. From August 2011 to April 2012, staff conducted 40 in-depth interviews with SHEWA-B recipients, seven focus group discussions with community members in SHEWA-B areas, and six interviews with key informants including school teachers, religious leaders, and local leaders. The interviews and focus groups were conducted in urban and rural SHEWA-B implementation areas in seven different regions of Bangladesh. To better understand factors affecting implementation quality, staff conducted additional in-depth interviews of 18 SHEWA-B recipients and six CHPs from June-July 2012. I classified sub-districts as high or low performing using initial cross-sectional survey results and conducted interviews in areas with high and low performance and a high and low percentage of target health behaviors. Interviews were recorded and transcribed in Bengali. Qualitative researchers at ICDDR,B manually coded the data and translated results into English and performed thematic content analysis.

Sample size calculation and sampling

The sample size was designed to be large enough to detect differences in outcomes between sub-districts. I assumed the design effect=2, alpha=0.05, power=0.8 and 28 observations per cluster and estimated mean outcomes using data from the 2009 SHEWA-B assessment. I calculated a required sample of 1,160 clusters in the 58 intervention sub-districts, yielding a total planned sample size of 32,480 households. The field team planned to interview 1,164 CHPs in 1,160 clusters (some CHPs were responsible for more than one cluster). The number of clusters per union (the geographic unit below an sub-district) was determined using probability proportionate to size (PPS) sampling based on the population size of each union. For each selected union, the number of village clusters was randomly selected. Villages that were previously sampled for other SHEWA-B assessments (e.g., diarrheal disease [39]) were excluded to minimize responding fatigue from repeated assessment. In selected clusters, field staff identified the center point of the village and used proximity sampling to select 28 households. Households with at least one child under five years were eligible for inclusion.

Implementation quality measurement

Since implementation quality was not defined a priori, I created an index using information from UNICEF about the factors they considered important predictors of CHP success.

ICDDR,B researchers employed the Delphi method to gather structured, qualitative feedback from 12 UNICEF staff that worked on SHEWA-B at the national headquarters and at the district level [40]. Each participant independently assigned points to variables from the cross-sectional survey and CHP survey that could measure implementation quality on a five-point scale (1=weak measure of implementation quality, 5=strong measure of implementation quality). Following the first round, researchers calculated the mean points per item, and reported the mean points per item individually to each participant, and asked participants if they wanted to change their initial point allocations. At each step participants gave qualitative feedback about the items. Researchers solicited suggestions for additional items in both rounds, and participants assigned points to these items.

We generated an implementation quality index using the average number of points in the second round for each item. I excluded items from the index that UNICEF staff did not consider useful and incorporated new items suggested by UNICEF if data was available. Each item included in the index received a weight equal to the mean points received. After reviewing the CHP survey responses, I chose to exclude most items from the CHP survey because I was concerned that CHPs had an incentive to report activities that they should have done rather than those they actually did. The only variable I retained from the CHP survey was recall of key SHEWA-B messages, which was less likely to be biased. I scaled the index so that the maximum value equaled 100 and the minimum equaled 0.

An index value of zero indicates that the respondent reported that they never met a SHEWA-B CHP nor ever heard about or attended a CHP community meeting and that the CHP in their community could not recall any of the key SHEWA-B messages. An index value of 100 indicates that the CHP visited the household in the last month and the respondent either heard of or attended a community event in the last month, knew the CHP's name, recalled that the CHP demonstrated key messages in the last year, and that the CHP could recall all key SHEWA-B messages.

Outcome definition and measurement

We measured selected target outcomes from UNICEF's log frame for SHEWA-B and several additional outcomes that could provide more objective measures of their targets. The outcomes were: 1) correct caregiver demonstration of handwashing (she used soap, water, both hands); 2) presence of a dedicated handwashing location within 10 feet of the place of defecation with water and soap (or if soap was not present, the respondent could retrieve soap within one minute); 3) observed clean child and caregiver hands (palms, finger pads, and fingernails were observed to be free of visible soil); 4) availability of a private, improved latrine according to the JMP (UNICEF/WHO) definition [41]; 5) observed no feces on the latrine slab or floor; 6) observed hygienic drinking water collection point (the platform at the water collection point was not broken or water logged, and there were no feces or garbage around it); 7) drinking water container was observed to be covered; 8) having received at least one water, sanitation, and hygiene promotion message from a SHEWA-B CHP; 9) and having ever received any water, sanitation, or hygiene promotion messages from a SHEWA-B CHP. With the exception of outcomes related to key messages, all outcomes were collected using spot checks.

Measurement of potential confounders

I pre-specified potential confounders as variables that could impact implementation quality and outcomes either directly or through intermediates. Using directed-acyclic-graphs, shown in Figure B.1, [42] I determined that sub-district-level poverty, the season of data collection, and geographic features (flood-prone and drought-prone areas) were potential confounders for all outcomes except for having an improved latrine, for which only sub-district-level poverty was a potential confounder. I used data from the 2000 Bangladesh Household Income and Expenditure Survey for the sub-district-level poverty variable [43]. Because NGOs recruited and trained CHPs, it is possible that in poorer areas, the education level of CHPs was lower or the training they received was of poorer quality. Lower education of respondents in poorer areas may also have affected the extent of behavior change. I defined cool season as September-February, hot season as March-May, and rainy season as June-August. In the rainy season, CHPs may have had more trouble traveling to assigned villages. I divided study areas into three geographic types: regular, haor area, and drought-prone areas. Haor areas are wetlands that are especially prone to flooding. CHPs may have had trouble traveling to villages in haor areas, particularly during the cool season, when paths becomes muddy. Households in drought-prone areas may have had limited access to water mainly in the dry season, which may have affected their handwashing and drinking water storage behaviors. I explored effect modification by each of these pre-specified, potential confounders for outcomes for which I considered effect modification to be plausible.

Analysis

We calculated summary statistics for outcomes and covariates and estimated 95% confidence intervals with robust standard errors adjusted for clustering at the village level [44]. I calculated the mean of target outcomes at the cluster-level within strata of sub-district poverty level, season, and geographic area and compared differences in means using a Wald test with standard errors adjusted for clustering.

In epidemiologic exposure analyses, a useful parameter of interest for measuring the health improvement attributable to either the removal or enhancement of an exposure is the population attributable risk and the related population attributable fraction, which quantify the change in health if an exposure were to be changed to a counterfactual distribution holding all other exposures at their observed values [45]. The population intervention parameter defined by Hubbard and van der Laan compares the prevalence of a disease at its counterfactual level to the current prevalence of disease in the population sample and estimates the population attributable risk within a causal inference model [46]. Such a parameter is useful for practitioners interested in understanding how an intervention could result in population-level changes given the current distribution of the intervention in the population. The population intervention parameter can also be used to assess the extent to which an intervention could have improved outcomes if implementation quality had been high compared to its observed level.

We measured the effect of implementation quality on outcomes by estimating a causal attributable risk (see Hubbard 2008 [46], for example) [47, 48]. Specifically, I estimated the difference in the mean probability of outcomes in clusters under the observed level of implementation quality and under a counterfactual scenario estimated from a model in which all clusters with an implementation quality index below the 75th percentile of the index were raised to that standard. I used a simple substitution estimator, which relies on a linear regression of outcomes versus the intervention of interest and confounders. I also fit the same model with a much more nonparametric procedure, using a data-adaptive, machine learning algorithm [49]. I used the following learners: generalized linear models, Bayesian main-terms logistic regression, lasso and elastic-net regularized GLM, generalized additive models, and stepwise regression with only main effect terms based on the Akaike Information Criterion. Results were similar using both estimation methods, so I only present the results from the standard linear regression using maximum likelihood. I also imputed the cluster-level mean probability of each outcome over values of the index. I then plotted the predicted values of the outcome across values of the index. I used a non-parametric bootstrap with 1,000 replicates to estimate standard errors and 95% confidence intervals for my estimates. Effect modification was considered to be statistically significant when at least two of the stratum-specific confidence intervals for point estimates did not overlap. To detect possible residual confounding of the association between implementation quality and target outcomes, I repeated the analysis using a negative control outcome: the number of neonatal deaths the respondent recalled in the last year [50].

3.3 Results

Sampling frame, response rate, and household characteristics

To reach the planned sample of 32,480 in 1,160 clusters, 33,134 households in 1,182 village clusters were invited to participate; 33,027 households consented to participate (response rate=99%), as shown in Table B.1. I attempted to reach all 1,164 CHPs, and were able to interview 1,110 CHPs in 1,126 clusters (95%). On average, 47% of respondents in a cluster reported that a CHP visited their households in the four months prior to the survey (n=1,126), and 26% of respondents in a cluster had heard of CHP-led community meetings (n=1,126). Table 3.1 summarizes the socio-demographic characteristics of respondents and sub-districts. Twenty-five percent of mothers and 38% of fathers had no education.

Table 3.1: Socio-demographic characteristics of upazilas, households, respondents.

	n	N	Percent/ mean
Male household head	30,307	31,465	96
Average household size	NA	31,465	5
Mother's education	NA	NA	
None	7,986	31,441	25
Up to primary	12,877	31,441	41
Up to secondary	10,340	31,441	33
Above secondary	238	31,441	1
Father's education	NA	NA	
None	11,997	31,284	38
Up to primary	10,201	31,284	33
Up to secondary	8,193	31,284	26
Above secondary	893	31,284	3
Proportion who own	NA	NA	
Electricity	15,635	31,416	50
Mobile phone	23,015	31,416	73
Television (B/W)	3,716	31,416	12
Television (color)	4,892	31,416	16
Refrigerator	1,537	31,416	5
Motorcycle	1,298	31,416	4
Home	29,633	31,445	94
Average amount of homestead land (acres)	NA	28,165	102
Average amount of land other than homestead (acres)	NA	15,054	170
Upazila-level poverty ∗	NA	NA	
0-24% poverty ∗	6,551	31,465	21
25-30% poverty ∗	4,900	31,465	16
31-36% poverty ∗	6,384	31,465	20
37-55% poverty	13,630	31,465	43
Season of data collection †	NA	NA	
Cool season	22,287	31,464	71
Hot season	2,273	31,464	7
Rainy season	6,904	31,464	22
Geographic features of upazila	NA	NA	
Drought-prone area	5,376	31,465	17
Haor area	2,884	31,465	9
Regular area	23,205	31,465	74
Average amount of homestead land (acres)	NA	28,165	102
Average amount of land other than homestead (acres)	NA	15,054	170

∗ Poverty incidence is defined as "the proportion of individuals living in that area who are in households with an average per capita expenditure below the (lower or upper) poverty line" according to a World Bank report using Bangladesh Bureau of Statistics data from 2001 (World Bank 2004).

† I defined cool season as September-February, hot season as March-May, and rainy season as June-August.

Forty-three percent of households sampled were in sub-districts in which 37-55% of residents were estimated to live below the poverty line. The majority of data (71%) was collected in the cool season and in non-haor areas not prone to drought (74%).

Features of the implementation quality index

Table 3.2 presents the percentage of households with each of the variables included in the implementation quality index. Forty-seven percent of households (n=14,622) reported meeting a SHEWA-B CHP at least once. Only 31% of households reported meeting a CHP in the last four months (n=8,328). Under half (47%) of CHPs surveyed could recall all three general key messages of SHEWA-B (n=561), and they were most likely to recall the message promoting handwashing with soap.

The first plot in Figure 3.2 shows the observed distribution of the implementation quality index. The mean index at the cluster level ranged from 0 to 90, with a mean of 28 and SD of 21 (n=1,126). The value of the index at the 75th percentile was 42– less than half the maximum possible value. The second plot in Figure 3.2 shows the distribution of the index under the counterfactual scenario.

Table 3.2: Inputs into implementation quality index.

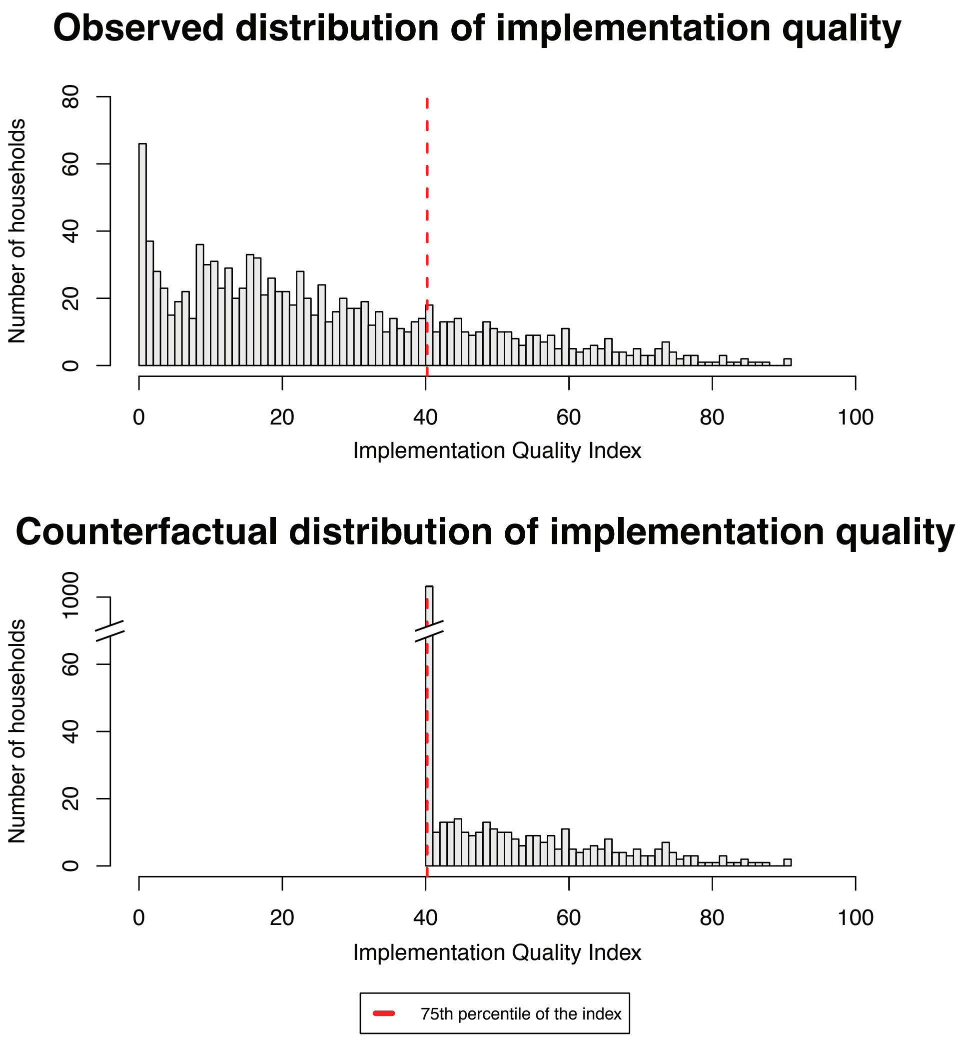
Input	n	Percent (95% CI)*
CHP visited household at least once	14,622	47 (45,48)
CHP visited household in the last month	5,471	19 (17,20)
CHP visited household in the last 4 months	8,328	31 (29,32)
Respondent ever heard of/attended a community event	7,892	26 (24,27)
Respondent heard of/attended a community event in the last month	3,341	11 (10,12)
Respondent heard of/attended a community event in the last quarter	5,821	19 (18,20)
Respondent knows a SHEWA-B CHP by name	8,663	28 (26,29)
Respondent recalls that a CHP gave safe water messages in the last year	8,076	26 (24,27)
Respondent recalls that a CHP gave handwashing messages in the last year	9,120	29 (28,31)
Respondent recalls that a CHP gave sanitation messages in the last year	9,128	29 (28,31)
Respondent recalls that a CHP gave at least 3 messages in the last year	7,584	24 (23,26)
CHP can recall all key SHEWA-B messages †	561	47 (44,50)

∗ Standard errors were adjusted for clustering at the cluster level.

† Data used to create this variable is from the CHP survey. If a cluster was covered by multiple CHPs, the values were averaged so that the N is the total number of clusters. The number presented in the "Percent" column is the mean across all clusters.

Table 3.3 presents the mean index values and 95% confidence intervals for potential confounders. The mean index value was lowest in the sub-districts with the lowest percentage of households under the poverty line. The mean was significantly higher in the cool and hot season than in the rainy season (30, 30, and 18, respectively). There was no statistically significant difference in the mean between different geographic areas. Figure 3.3 illustrates how to interpret the index values at 0, the 75th percentile (a score of 42), and 100 using four of the variables included in the index.

Figure 3.2: Histogram of the implementation quality index



Outcomes stratified by whether respondent ever met a CHP

Because such a large proportion of respondents never met a SHEWA-B CHP, it is possible that outcomes would be closer to UNICEF targets among those who had ever met a SHEWA-B CHP. In Table 3.5, I stratified the percentage of respondents with each outcome by whether they reported ever meeting a SHEWA-B CHP and compared the percentages using a Wald test. The percentage was significantly higher except for access to a private improved latrine, having no feces present on the latrine slab or floor, observed hand cleanliness, and having a hygienic drinking water point; however, note that even quite small differences that may not be programmatically meaningful are statistically significant due to the large sample size. The similarity in target outcome percentages among those who did and did not ever meet a CHP suggests that these outcomes were not sensitive to the CHP intervention. I also explored whether the probability of outcomes increased with the frequency of CHP visits; there was an increasing pattern for correct caregiver handwashing and having received water, sanitation, and hygiene messages from a SHEWA-B CHP (see Table B.2).

Table 3.3: Index mean stratified by covariates.

Covariate	n	Mean (95% CI)*
0-24% poverty †	6,551	21 (19,23)
25-30% poverty †	4,900	31 (28,35)
31-36% poverty †	6,384	33 (30,36)
37-55% poverty †	13,630	27 (25,29)
Cool season	22,287	31 (29,32)
Hot season	2,273	30 (26,34)
Rainy season	6,904	18 (15,20)
Drought-prone area	5,376	26 (26,29)
Haor area	2,884	31 (26,36)
Regular area	23,205	28 (23,28)

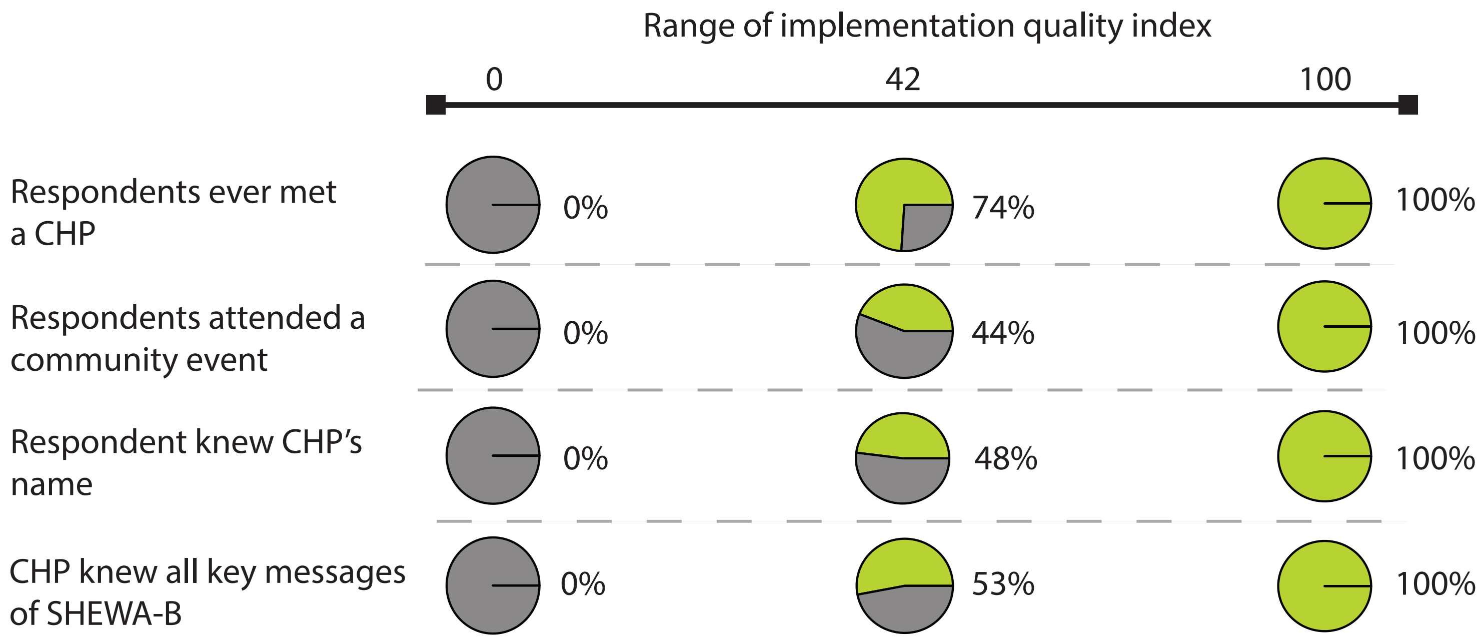
∗ Standard errors were adjusted for clustering at the cluster level.

† Poverty incidence is defined as "the proportion of individuals living in that area who are in households with an average per capita expenditure below the (lower or upper) poverty line" according to a World Bank report using Bangladesh Bureau of Statistics data from 2001 (World Bank 2004).

Outcomes compared to UNICEF targets

Performance was close to UNICEF targets for 1) having no feces on the latrine slab or floor (observed: 50%, target: 59%), 2) presence of a dedicated handwashing location (observed: 57%, target: 55%), 3) no open defecation (observed: 94%, target: 97%), and 4) covering drinking water containers (observed: 43%, target: 45%). Performance was substantially below target for 1) access to a private, improved latrine (observed: 23, target: 75%), 2) having a hygienic drinking water point (observed: 28%, target: 82%), and 3) receiving messages from a SHEWA-B CHP (observed: 45%, target: 82%). These results are displayed in Table 3.4.

Figure 3.3: Heuristic guide to interpreting the implementation quality index



This figure illustrates the typical implementation characteristics at different levels of the implementation quality index. The pie charts display the observed percentage of respondents with each variable in clusters in which the index was equal to 0 or 42 (the 75th percentile) and hypothetical values when the index was 100 (since there were no clusters which had an index value equal=100). To interpret the first row of the figure, in clusters with index=0, no respondents had ever met a CHP. In clusters at the 75th percentile (index=42), 74% of respondents ever met a CHP. If there had been clusters with index=100, all respondents would have ever met a CHP. This figure contains four of twelve variables included in the implementation quality index for simplicity.

Table 3.4: Outcomes and UNICEF endline targets

Outcome	n	N*	Percent (95% CI)†	UNICEF endline target
Private, improved latrine available	6,725	29,586	23 (22,24)	75
No feces on latrine slab or floor ‡	14,284	28,360	50 (49,51)	59
Handwashing station available	17,988	31,431	57 (56,59)	55
No open defecation	29,523	31,449	94 (93,94)	97
Correct caregiver handwashing	18,096	29,894	61 (59,62)	
Caregiver hands observed to be clean §	13,587	31,291	43 (42,44)	
Child hands observed to be clean §	8,455	29,340	29 (28,30)	
Received at least 1 W,S,H message from CHP ||	10,948	31,170	35 (44,47)	82
Respondent heard any messages from a SHEWA-B CHP ||	14,169	31,254	45 (34,37)	82
Drinking water point is sanitary ¶	8,801	31,312	28 (27,29)	82
Drinking water container covered	5,305	12,253	43 (42,45)	45

*Although the index is calculated at the cluster level, outcomes are calculated at the household level, so the N is shown at the household level. 1,178 clusters had non-missing values for the index.

† Standard errors were adjusted for clustering at the cluster level.

‡ In the UNICEF log frame, this indicator is for the percent of rural latrines, but I have estimated it as the percent of rural households with latrines.

§ No visible presence of dirt on nails, palms or finger pads

|| The UNICEF log frame corresponding to these items was somewhat general, so these variables may estimate something slightly different than what UNICEF intended.

¶ Environmental sanitation is considered maintained if the water point's platform is not broken and not water logged and has no garbage, dirt, or feces around it.

Table 3.5: Outcomes stratified by whether respondent ever met a CHP

Outcome	N	% if never met CHP	% if ever met CHP	p-value
Private, improved latrine available	29,434	23	23	0.609
No feces on latrine slab or floor	28,212	51	50	0.412
No open defecation	31,288	93	95	0.001
Has dedicated handwashing location	31,271	55	60	0.000
Correct caregiver handwashing	29,746	57	65	0.000
Caregiver hands observed to be clean	31,134	44	43	0.304
Child hands observed to be clean	29,217	29	29	0.552
Received at least 1 W,S,H message from CHP	31,035	0	76	0.000
Respondent heard any messages from a SHEWA-B CHP	31,155	0	97	0.000
Drinking water container covered	12,209	42	46	0.001
Drinking water point is sanitary	31,146	30	26	0.000

Effect of implementation quality on outcomes

Figure 3.4 shows the estimated effect the SHEWA-B intervention would have had for each outcome if it had been implemented at the 75th percentile or higher in all clusters compared to the effect of the intervention given the observed level of implementation quality. I present stratified estimates for outcomes for which effect modification was present; otherwise main effects are presented. No open defecation was modified by sub-district-level poverty and season, and clean caregiver hands were modified by season. There was a significant increase in the probability of no open defecation under the increased implementation quality scenario in the cool season, no statistically significant difference in the hot season, and a decrease in the rainy season. Clean caregiver hands and correct caregiver handwashing had the largest effect sizes; the probability of correct caregiver handwashing increased 6.0% (95% CI 4.5%, 7.5%) in households living in sub-districts with 37-55% of households below the poverty line when improving the quality of implementation of the intervention to at least the 75th percentile of the index in all households. The effect size was smaller for households in relatively wealthier sub-districts. Interestingly, the probability of clean caregiver hands increased the most for households measured in the rainy season (4.2%; 95% CI 2.1%, 6.6%) and decreased in the other seasons. For the majority of outcomes, increases in the probability of the outcome associated with improved implementation quality were less than 2 percentage points, indicating that increased implementation quality was not associated with improved outcomes (see Table B.3). The point estimate for the negative control outcome of neonatal deaths in the past year was 0.0008, indicating no association between this variable and the index.

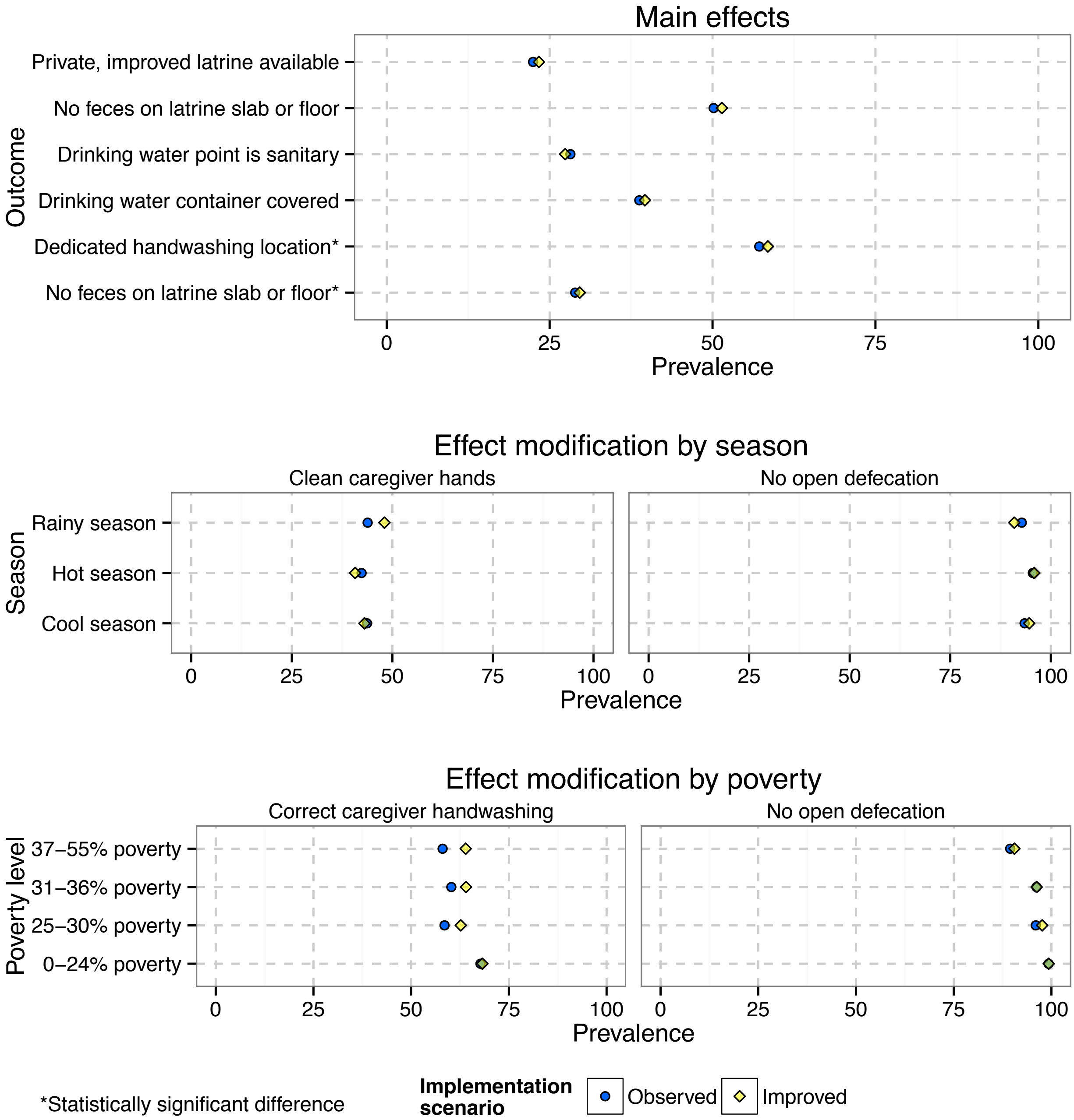
Dose-response relationship between implementation quality and outcomes

We explored the potential for a dose response relationship between implementation quality and the probability of each outcome. Figure 3.5 shows the observed and predicted values for having a private, improved latrine by the implementation quality index. The predicted values from the regression model suggest a miniscule increase in the probability of the outcome as the index increases but no clear dose response pattern. The observed data is not clustered tightly around the smoothed line from the regression. This pattern was similar for all outcomes.

CHP survey results

The field team asked CHPs an open ended question about problems in their work; the most commonly reported problems were that beneficiaries did not have time to listen during community meetings (n=495, 45%) and household visits (n=376, 34%), that beneficiaries did not have enough money to buy soap (n=375, 34%) and that beneficiaries were not interested in attending community meetings (n=365, 33%). The majority reported that they met with

Figure 3.4: Prevalence of outcomes under observed implementation quality and under counterfactual scenario



their supervisor at least weekly (n=814, 73%) and that supervision was sufficient (n=908, 82%). The majority also reported satisfaction with the content (n=934, 84%) and duration of their training (n=709, 64%). Only 72 (6%) CHPs reported that they had other jobs. The majority reported that their stipend was insufficient (n=961, 87%) and was not paid on time (n=721; 65%). When asked to recall specific key messages of SHEWA-B, on average, CHPs recalled 2.9 out of 5 messages about safe water storage, 4.5 out of 6 about handwashing, and 4.5 out of 9 about latrine usage.

Findings of in-depth interviews and focus groups

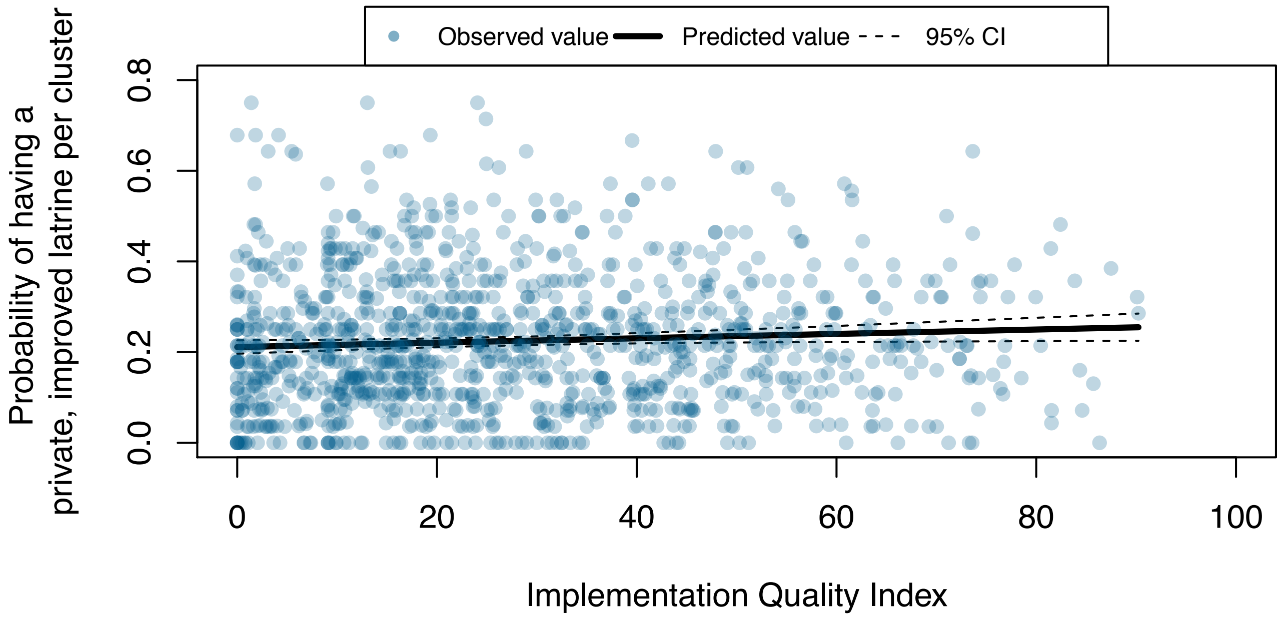
All people invited to participate in the August 2011 - April 2012 focus groups and interviews agreed to participate. Of the 25 people invited to participate in the June-July 2012 in-depth interviews, one declined. Table 3.6 contains the demographic characteristics of respondents. Half of SHEWA-B recipients had at least a secondary level education, and more than half considered themselves to have a low income.

Barriers to intervention uptake

SHEWA-B recipients reported scarce physical resources and limited land on which to install improved latrines. In areas with poor health outcomes, almost all participants reported limited resources to buy soap, limited time, lethargy, and tradition as barriers to handwashing. SHEWA-B recipients reported limited available land to install tubewells, limited availability of safe drinking water in the dry and rainy seasons, and having to walk long distances and wait a long time to collect water from shared water sources. Respondents mentioned insufficient promotion of behavior change by CHPs and disinterest in CHPs. In low performing areas, the majority of respondents reported that CHPs only delivered SHEWA-B messages and never followed up with participants. The majority also reported a lack of willingness give time to CHPs, unwillingness to follow CHP suggestions, and unwillingness to attend CHP meetings.

In low performing areas, 11 out of 18 SHEWA-B recipients reported that they did not see any local leaders attend the CHP-led community meetings. Two CHPs reported that local leaders were not interested because they did not receive financial remuneration and were more interested in hardware distribution programs. In low CHP performance areas, respondents reported that another large non-governmental organization (NGO) had provided improved latrines with easy credit installments there since mid-2010. A female CHP from Comilla said, “A big NGO is providing free latrines to its beneficiaries. But we do not have provision to distribute any hardware support like them. Do you think people will listen to us if we do not give them anything except words?”

Figure 3.5: Example of potential dose-response relationship



Motivators for intervention uptake

The majority (10 out of 18) of SHEWA-B recipients reported attending community meetings and that performances with songs and pictures were effective drivers of behavior change. In high performing areas, six out of eight participants reported that CHPs visit their households regularly, reminded them about hygiene promotion, inspected latrine cleanliness, and made suggestions about how to solve any problems. In these areas, CHPs reported that community leader involvement contributed to their success.

Barriers to CHP performance

In all three areas, almost every CHP reported that their stipend was too low and was often paid irregularly. A male CHP from Panchagar said, "We receive only 98 taka [approximately 1.27 USD] per day which is less than half of a daily laborer's payment. If I were not unemployed, I would not do this job at all." Focus group respondents reported that political influences affected about 20-30% of the CHP hires and that CHPs hired in this fashion showed less interest in their job duties. In contrast with the results of the CHP survey, focus group respondents reported that over a quarter of CHPs were involved in other work, such as business, or were full time students.

Table 3.6: Demographic characteristics of focus group and in-depth interview respondents

	August 2011-April 2012 Assessment		June-July 2012 Assessment	
Demographic characteristics	In-depth interviews with key informants (n=40)	Focus group discussions with SHEWA-B recipients (n=40)	In-depth interviews with SHEWA-B recipients (n=18)	In-depth interviews with CHPs (n=6)
Gender				
Male	0	28	0	1
Female	40	12	18	5
Age				
18 - 23 yrs	9	7	7	2
24 - 29 yrs	16	4	6	2
30 - 35 yrs	11	6	5	2
36 - 41 yrs	3	6	0	0
42 - 59 yrs	1	8	0	0
60 - 65 yrs	0	9	0	0
Education				
No education	14	8	3	0
Primary level	22	27	6	0
Secondary level	4	3	8	0
Higher secondary	0	0	0	4
Over higher secondary	0	2	1	2
Self-reported economic status				
Poor	16	11	6	0
Lower middle class	13	11	4	4
Middle class	0	0	4	2
Upper middle class	11	13	2	0
Rich	0	5	2	0

3.4 Discussion

This large-scale (N=33,027 households), population-based assessment of SHEWA-B is among the largest assessments ever conducted of a WASH program. Delivery of SHEWA-B was sub-optimal: the majority of respondents did not recall ever meeting a CHP. Low exposure to CHPs was the main factor driving observed suboptimal implementation quality as assessed by the quality index. Outcomes were only marginally better among households who had met a CHP. Despite the reports of successful performance of individual CHPs in some areas from the qualitative assessment, the observed distribution of implementation quality suggests that implementation quality did not meet UNICEF's ideal in any area. Although some outcomes were close to UNICEF targets, the small associations between implementation quality and outcomes suggest that observed health behaviors may be better attributed to factors outside of SHEWA-B than to SHEWA-B itself. For instance, sub-district-level poverty was a stronger predictor of access to a private, improved latrine than how frequently a CHP visited. These modest findings highlight the difficulty of maintaining high quality implementation at scale. The forthcoming endline SHEWA-B assessment results include a control group and should reveal whether outcomes improved concurrently outside of SHEWA-B areas during the intervention.

Our implementation quality index was developed through a systematic process with UNICEF staff who designed and implemented SHEWA-B, but it remains possible that it was poorly defined. The credibility of my findings are strengthened because I found no association between the index and a negative control outcome (neonatal deaths), suggesting that the association I report between implementation quality and target outcomes was not likely to be a result caused by residual confounding. Future studies of large-scale interventions would benefit from concise, a priori definition of the intervention and intervention fidelity measures to allow for rigorous, generalizable assessment [51, 52]. Because the observed range of the index did not reach the maximum possible value, I was unable to estimate the effects under a scenario in which all households received a perfectly implemented intervention. If more data had been available for higher values of the index, I could have defined the counterfactual scenario at, for example, the 90th percentile rather than the 75th, and I may have observed a larger effect size. Given the available, observed data, doing so would have relied on a model to extrapolate beyond the information in the observed data and would be prone to bias.

It is possible that recall bias, respondent bias, and measurement error occurred. The field team used rapid observations of hygiene practices and conditions because they are efficient and have been shown to be valid, reliable indicators for many hygiene outcomes [53]. I also augmented UNICEF's target outcome list with additional outcomes which have been shown to be less biased, such as observed hand cleanliness [53]. A major limitation was that CHP survey responses were not consistent with the findings of the cross-sectional survey and qualitative assessment, suggesting considerable response bias from CHPs. Such bias, though not surprising, highlights the difficulty of evaluating CHPs in large-scale interventions and the value of qualitative data.

There are a number of ways in which the SHEWA-B intervention could have been better designed. The CHP literature suggests that sufficient supervision and remuneration contribute to CHP success [21, 54, 55]. The number of households SHEWA-B CHPs were responsible for may have been unreasonable (1200-1500 per CHP). Even if the workload was manageable, there was likely a high opportunity cost, particularly considering that the majority of CHPs considered their stipend to be insufficient [54], and focus group respondents reported that some CHPs had other jobs. Other international non-governmental organizations working in rural Bangladesh paid similar types of workers approximately 2.50 USD per day at that time. Although CHPs reported satisfaction with their training, it is possible that the training was insufficient. Indeed, CHPs appeared to need more training: on average, CHPs could only recall 2.9 out of 5 messages about safe water storage, 4.5 out of 6 about handwashing, and 4.5 out of 9 about latrine usage. The intervention may have been more successful if SHEWA-B had provided facilitative hardware, such as handwashing devices, to each participating household in addition to health promotion messages, as other local NGOs did [56]. See Appendix 2 for a list of all hardware provided by SHEWA-B. There is a growing body of evidence suggesting that hygiene behavior in low-income countries does not change in response to health education but rather to other factors, such as social acceptance and disgust with feces [57–59]. Researchers are also exploring whether subsidizing water, sanitation, and hygiene hardware results in higher uptake than offering free or market-rate hardware, although there are concerns that such approaches fail to reach the poorest of the poor, who were targeted by SHEWA-B [60–62].

Even though many CHPs reported sufficient supervision, inadequate supervision may have contributed to suboptimal CHP performance. Some studies found that improved supervision and audits with feedback improved CHP performance and increased CHP job satisfaction and motivation [21]. Although UNICEF and DPHE conducted performance assessments of CHPs and higher-level staff, the results suggest that assessments did not result in high level CHP performance in most areas. Political influence during the process of hiring CHPs may have also indirectly contributed to suboptimal CHP performance, as has been noted in the CHP literature [54]. Focus group respondents mentioned that NGOs were subject to political and social influences and that in these cases CHPs showed less interest in their job duties. As has been noted in the literature, in future CHP interventions, a higher salary, improved supervision and training, more thorough assessment, and a more transparent hiring process would likely contribute to improved CHP performance [21, 54, 55].

The results of this large assessment demonstrate the difficulty of maintaining intervention quality while expanding coverage on a large scale, as has been reported by others [13]. On the whole, these findings echo those of the few, existing assessments of large-scale WASH interventions: in Indonesia, Peru, and Vietnam found no effect on access to improved sanitation and mixed results related to handwashing and diarrhea prevalence [35–37]. A large-scale sanitation intervention in India was found to greatly increase sanitation coverage but did not reduce disease [63].

The literature on scaling up has reported that factors for successful interventions at scale include strong leadership and management, realistic arrangements for financing, country ownership of the intervention, and technical innovation [13]. As such, when interventions are similar at the household or community level, differences in the impact of scaled interventions between countries may be explained by higher level factors such as governance. I was unable to explore the role of factors such as management and financing of SHEWA-B because of the complexity of implementation. Implementing organizations could conduct further qualitative research to improve their understanding of how these factors might have affected implementation quality and outcomes.

These results illustrate the use of rigorous methods to systematically evaluate implementation quality for programs at-scale and demonstrates their potential value for improving and refining program delivery. Rigorous assessment of WASH interventions is difficult, even for small-scale interventions [64], and very few large-scale WASH interventions have been evaluated rigorously [65]. In particular, assessment of implementation can be difficult when not built into the assessment from its inception. I evaluated the impact of improved implementation of the SHEWA-B program by employing statistical methods developed in the causal inference literature as well as qualitative methods. My approach using population intervention models has broad applicability to similar assessments of other large-scale public health programs. These findings contribute to the growing empirical evidence base describing best practices for and barriers to delivering interventions at scale. Such evidence may contribute to improvements in design and delivery of interventions which in turn could increase the health impact of such interventions when delivered at scale [13].

Bibliography

[1] *World Development Report 1993: Investing in Health*. World Bank and Oxford University Press, 1993.

[2] Sachs JD. Health in the developing world: achieving the Millennium Development Goals. *Bulletin of the World Health Organization* 82:12 (2004), pp. 947–949.

[3] *Scaling Up: From Vision to Large-Scale Change*. Management Systems International, 2012. URL: `http://www.msiworldwide.com/wp-content/uploads/Scaling-Up-Framework.pdf`.

[4] Duflo E. *Scaling Up and Evaluation*. 2004. URL: `http://economics.mit.edu/files/766`.

[5] Bryce J, Black RE, Walker N, Bhutta ZA, Lawn JE, and Steketee RW. Can the world afford to save the lives of 6 million children each year? *The Lancet* 365:9478 (2005), pp. 2193–2200.

[6] Jha P, Mills A, Hanson K, Kumaranayake L, Conteh L, Kurowski C, Nguyen SN, Cruz VO, Ranson K, Vaz LM, et al. Improving the health of the global poor. *Science* 295:5562 (2002), pp. 2036–2039.

[7] Victora CG, Hanson K, Bryce J, and Vaughan JP. Achieving universal coverage with health interventions. *The Lancet* 364:9444 (2004), pp. 1541–1548.

[8] Bloom G and Ainsworth P. *Beyond scaling up: pathways to universal access to health services*. STEPS Working Paper 40. STEPS Centre, 2010.

[9] Tangcharoensathien V and Patcharanarumol W. Commentary: Global health initiatives: opportunities or challenges? *Health Policy and Planning* 25:2 (2010), pp. 101–103.

[10] Lu C, Schneider MT, Gubbins P, Leach-Kemon K, Jamison D, and Murray CJ. Public financing of health in developing countries: a cross-national systematic analysis. *The Lancet* 375:9723 (2010), pp. 1375–1387.

[11] Madon T, Hofman K, Kupfer L, and Glass R. Implementation Science. *Science* 318:5857 (2007), pp. 1728–1729.

[12] McCannon CJ, Berwick DM, and Massoud MR. The science of large-scale change in global health. *JAMA* 298:16 (2007), pp. 1937–1939.

[13] Mangham LJ and Hanson K. Scaling up in international health: what are the key issues? *Health Policy and Planning* 25:2 (2010), pp. 85–96.

[14] Sanders D and Haines A. Implementation research is needed to achieve international health goals. *PLoS Medicine* 3:6 (2006), e186.

[15] Subramanian S, Naimoli J, Matsubayashi T, and Peters DH. Do we have the right models for scaling up health services to achieve the Millennium Development Goals? *BMC Health Services Research* 11:1 (2011), p. 336.

[16] Gilson L, Walt G, Heggenhougen K, Owuor-Omondi L, Perera M, Ross D, and Salazar L. National community health worker programs: how can they be strengthened? *Journal of public health policy* (1989), pp. 518–532.

[17] Chopra M and Ford N. Scaling up health promotion interventions in the era of HIV/AIDS: challenges for a rights based approach. *Health Promotion International* 20:4 (2005), pp. 383–390.

[18] Durlak JA. Why program implementation is important. *Journal of Prevention & Intervention in the community* 17:2 (1998), pp. 5–18.

[19] Durlak JA and DuPre EP. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology* 41:3-4 (2008), pp. 327–350.

[20] Hanson K, Ranson MK, Oliveira-Cruz V, and Mills A. Expanding access to priority health interventions: a framework for understanding the constraints to scaling-up. *Journal of International Development* 15:1 (2003), pp. 1–14.

[21] Rowe AK, Savigny D de, Lanata CF, and Victora CG. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *The Lancet* 366:9490 (2005), pp. 1026–1035.

[22] Yamey G. Scaling up global health interventions: a proposed framework for success. *PLoS Medicine* 8:6 (2011), e1001049.

[23] Clasen TF, Bostoen K, Schmidt WP, Boisson S, Fung I, Jenkins MW, Scott B, Sugden S, and Cairncross S. Interventions to improve disposal of human excreta for preventing diarrhoea. *Cochrane Database of Systematic Reviews* 6 (2010).

[24] Clasen T, Schmidt WP, Rabie T, Roberts I, and Cairncross S. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *BMJ* 334:782 (2007).

[25] Curtis V and Cairncross S. Effect of washing hands with soap on diarrhoea risk in the community: a systematic review. *The Lancet Infectious Diseases* 3:5 (2003), pp. 275–281.

[26] Daniels DL, Cousens SN, Makoae L, and Feachem RG. A case-control study of the impact of improved sanitation on diarrhoea morbidity in Lesotho. *Bulletin of the World Health Organization* 68:4 (1990), pp. 455–463.

[27] Doocy S and Burnham G. Point-of-use water treatment and diarrhoea reduction in the emergency context: an effectiveness trial in Liberia. *Tropical Medicine & International Health* 11:10 (2006), pp. 1542–1552.

[28] Ejemot-Nwadiaro R, Ehiri J, Meremikwu M, and Critchley J. Hand washing for preventing diarrhoea. *Cochrane Database of Systematic Reviews* 1 (2008).

[29] Fewtrell L, Kaufmann RB, Kay D, Enanoria W, Haller L, and Colford Jr JM. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *The Lancet Infectious Diseases* 5:1 (2005), pp. 42–52.

[30] Luby SP, Agboatwalla M, Feikin DR, Painter J, Billhimer W, Altaf A, and Hoekstra RM. Effect of handwashing on child health: a randomised controlled trial. *The Lancet* 366:9481 (2005), pp. 225–233.

[31] Rabie T and Curtis V. Handwashing and risk of respiratory infections: a quantitative systematic review. *Tropical Medicine & International Health* 11:3 (2006), pp. 258–267.

[32] Roberts L, Chartier Y, Chartier O, Malenga G, Toole M, and Rodka H. Keeping clean water clean in a Malawi refugee camp: a randomized intervention trial. *Bulletin of the World Health Organization* 79:4 (2001), pp. 280–287.

[33] Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, Jha P, Campbell H, Walker CF, Cibulskis R, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *The Lancet* 375:9730 (2010), pp. 1969–1987.

[34] Prüss-Üstün A, Bos R, Gore F, and Bartram J. *Safer water, better health: costs, benefits and sustainability of interventions to protect and promote health.* World Health Organization, 2008.

[35] Cameron L, Shah M, and Olivia S. *Impact evaluation of a large-scale rural sanitation project in Indonesia.* World Bank, Washington, DC, 2013.

[36] Chase C and Do QT. *Scaling Up Handwashing Behavior: Findings from the Impact Evaluation Baseline Survey in Vietnam.* World Bank, Washington, DC, 2010.

[37] Galiani S, Gertler P, and Orsola-Vidal A. *Promoting Handwashing Behavior in Peru: The Effect of Large-Scale Mass-Media and Community Level Interventions. SSRN Scholarly Paper.* 2012. URL: http://papers.ssrn.com/abstract=2170640.

[38] Feachem RG. Editorial: the research imperative: fighting AIDS, TB and malaria. *Tropical Medicine & International Health* 9:11 (2004), pp. 1139–1141.

[39] Huda TMN, Unicomb L, Johnston RB, Halder AK, Yushuf Sharker MA, and Luby SP. Interim evaluation of a large scale sanitation, hygiene and water improvement programme on childhood diarrhea and respiratory disease in rural Bangladesh. *Social Science & Medicine* 75:4 (2012), pp. 604–611.

[40] Linstone HA and Turoff M. The Delphi method: Techniques and applications (1975).

[41] WHO and UNICEF. *Types of drinking-water sources and sanitation*. 2013. URL: http://www.wssinfo.org/definitions-methods/watsan-categories/.

[42] Greenland S, Pearl J, and Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 10:1 (1999), pp. 37–48.

[43] Bangladesh Bureau of Statistics. Household Income and Expenditure Survey (2000). URL: http://www.bbs.gov.bd/PageReportLists.aspx?PARENTKEY=66.

[44] Huber PJ. "The behavior of maximum likelihood estimates under nonstandard conditions". *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 1. 1967, pp. 221–233.

[45] Rothman KJ, Greenland S, and Lash TL. *Modern epidemiology*. Lippincott Williams & Wilkins, 2008.

[46] Hubbard A and van der Laan M. Population Intervention Models in Causal Inference. *Biometrika* 95 (2008), pp. 35–47.

[47] Muñoz ID and van der Laan M. Population intervention causal effects based on stochastic interventions. *Biometrics* 68:2 (2012), pp. 541–549.

[48] Greenland S and Drescher K. Maximum Likelihood Estimation of the Attributable Fraction from Logistic Models. *Biometrics* 49 (1993), pp. 865–872.

[49] van der Laan MJ, Polley EC, et al. Super learner. *Statistical Applications in Genetics and Molecular Biology* 6:1 (2007), pp. 1–23.

[50] Lipsitch M, Tchetgen ET, and Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology (Cambridge, Mass.)* 21:3 (2010), p. 383.

[51] Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, and Tyrer P. Framework for design and evaluation of complex interventions to improve health. *BMJ: British Medical Journal* 321:7262 (2000), p. 694.

[52] Hawe P, Shiell A, Riley T, and Gold L. Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. *Journal of Epidemiology and Community Health* 58:9 (2004), pp. 788–793.

[53] Ram P. Practical guidance for measuring handwashing behavior. *World Bank Water and Sanitation Programme, Washington, DC, USA http://www. wsp. org/sites/wsp. org/files/PracticalGuidance_HWWS. pdf (date of visit: 23.05. 2013)* (2010).

[54] Haines A, Sanders D, Lehmann U, Rowe AK, Lawn JE, Jan S, Walker DG, and Bhutta Z. Achieving child survival goals: potential contribution of community health workers. *The Lancet* 369:9579 (2007), pp. 2121–2131.

[55] Swider SM. Outcome effectiveness of community health workers: an integrative literature review. *Public Health Nursing* 19:1 (2002), pp. 11–20.

[56] Dreibelbis R, Winch PJ, Leontsini E, Hulland KR, Ram PK, Unicomb L, and Luby SP. The integrated behavioural model for water, sanitation, and hygiene: a systematic review of behavioural models and a framework for designing and evaluating behaviour change interventions in infrastructure-restricted settings. *BMC Public Health* 13:1 (2013), p. 1015.

[57] Scott B, Curtis V, Rabie T, and Garbrah-Aidoo N. Health in our hands, but not in our heads: understanding hygiene motivation in Ghana. *Health Policy and Planning* 22:4 (2007), pp. 225–233.

[58] Curtis VA, Danquah LO, and Aunger RV. Planned, motivated and habitual hygiene behaviour: an eleven country review. *Health Education Research* 24:4 (2009), pp. 655–673.

[59] Biran A, Schmidt WP, Wright R, Jones T, Seshadri M, Isaac P, Nathan N, Hall P, McKenna J, Granger S, et al. The effect of a soap promotion and hygiene education campaign on handwashing behaviour in rural India: a cluster randomised trial. *Tropical Medicine & International Health* 14:10 (2009), pp. 1303–1314.

[60] Zwane AP and Kremer M. What works in fighting diarrheal diseases in developing countries? A critical review. *The World Bank Research Observer* 22:1 (2007), pp. 1–24.

[61] Curtis VA, Garbrah-Aidoo N, and Scott B. Ethics in public health research: masters of marketing: bringing private sector skills to public health partnerships. *American Journal of Public Health* 97:4 (2007), p. 634.

[62] Rheingans R, Dreibelbis R, and Freeman M. Beyond the Millennium Development Goals: public health challenges in water and sanitation. *Global Public Health* 1:1 (2006), pp. 31–48.

[63] Patil SR, Arnold BF, Salvatore A, Briceno B, Colford Jr JM, and Gertler PJ. A randomized, controlled study of a rural sanitation behavior change program in Madhya Pradesh, India. *World Bank Policy Research Working Paper* 6702 (2013).

[64] Blum D and Feachem RG. Measuring the impact of water supply and sanitation investments on diarrhoeal diseases: problems of methodology. *International Journal of Epidemiology* 12:3 (1983), pp. 357–365.

[65] Cairncross S, Bartram J, Cumming O, and Brocklehurst C. Hygiene, sanitation, and water: what needs to be done? *PLoS Medicine* 7:11 (2010).

Chapter 4

Advances in the estimation of the population attributable fraction: application of a causal inference technique to simulated and empirical datasets

4.1 Introduction

Randomized trials have long been considered the gold standard for evaluating community-based public health interventions in epidemiology. More recently, related fields, such as economics, have increasingly used trials [1]. While trials are often highly internally valid, they are also often subject to limitations that restrict their generalizability and utility for important research questions in public health. When one's goal is to evaluate the effectiveness of a large-scale intervention that is known to be efficacious in ideal settings, it is often neither feasible nor ethical to randomize [2]. Because randomization necessitates delivery of an intervention and clear differentiation between intervention and control groups, the range of exposure measurement in trials is often limited. Measurement is also often limited to a few key outcomes because research questions are narrowly focused. The contrasts between exposure states are limited by the design of the trial. Poor compliance to trial interventions can limit inference, and reasons for poor compliance in a trial may differ from in other settings, limiting generalizability. In addition, the populations enrolled in trials are often not representative because they are selected to answer a focused research question or to ensure high compliance. For these reasons, observational designs are an alternative that should not be overlooked. While subject to many pitfalls of their own, observational studies can yield highly relevant, generalizable findings, particularly when they are designed and analyzed thoughtfully.

For any study design, careful definition of the quantity to be estimated (i.e. the parameter) is critical to ensuring that the analysis estimates a quantity that helps answer the research question. When one is interested in understanding the population-level impact of an intervention the population attributable fraction (PAF) is often an appropriate parameter [3]. A variety of formulations of the population attributable fraction have been proposed; here I specifically refer to the parameter defined by Levin (1953), which compares the observed probability of an outcome (Y) to the probability under a counterfactual scenario in which an exposure (A) was removed (PAF=$(P(Y) - P(Y|A = 0))/P(Y)$) [4]. Bruzzi, Greenland, Drescher, and others have described model-based methods for estimating population attributable fractions that adjust for covariates and allow for continuous exposures [5–7].

Morgenstern and Bursic (1982) proposed the generalized impact fraction (GIF), which extended the PAF to allow the counterfactual to be defined in a variety of ways [8]. They defined the parameter as GIF=$(P(Y) - P(Y^*))/P(Y)$, where $P(Y^*)$ denotes the probability of Y under a modified distribution of exposure. The counterfactual (Y^*) can be defined such that an exposure or intervention is modified or reduced rather than eliminated, and it can be estimated for binary, categorical, or continuous exposures. More recently, Hubbard and van der Laan proposed the population intervention model (PIM), which is akin to the generalized impact fraction in that it allows the investigator to tailor the definition of the parameter to answer particular research questions. The PIM parameter compares the observed mean in the population to the mean outcome under a counterfactual scenario of one's choice [9]. The parameter is defined as $E[Y_{A^*} - Y]$, where Y_{A^*} denotes the counterfactual outcome under a modified treatment, and frequently treatment is modified to be improved compared to the observed level of treatment. For example, if one is interested in estimating the proportion of global mortality that could be prevented by water and sanitation interventions, one could compare the mortality under the current global setting, in which a large portion of the world does not have access to functional piped water and sewerage, compared to an ideal counterfactual setting in which over 90% of people in each country had access. PIM parameters can be defined as a single quantity comparing mortality under these two scenarios.

Most parameters used in epidemiology assume that interventions are deterministic (i.e., that receiving the intervention always results in a particular outcome). However, one can also define parameters that allow for stochastic interventions or counterfactuals, which assume that those receiving an intervention have a particular probability of the outcome but that the outcome is not guaranteed to occur. The distinction between stochastic and deterministic interventions is especially important when studying interventions that cannot be controlled or manipulated by the investigator, and thus it is not realistic to assume that they are deterministic. For example, if policy was implemented to reduce levels of air pollution with the goal of reducing asthma attacks, a deterministic policy would always result in a reduction in asthma attacks, while in a stochastic program, some individuals would have fewer asthma attacks and others would not. It is much more reasonable to assume a stochastic intervention for this particular example since a policy to decrease air pollution might not yield equivalent air pollution reductions in all areas and because individuals' responses to air pollution reductions might be a function of current health status, immunity, personal level of exposures, and other factors. When interventions are concerned with social and behavior phenomena, such as health education or handwashing promotion interventions, assuming that interventions are stochastic is generally more appropriate. Responses to interventions can also be stochastic in studies of biologic measures; for example, Cain et al. defined stochastic counterfactuals in a study comparing outcomes when HIV treatment was initiated after CD-4 cell counts drop below different thresholds [10]. Deterministic interventions can be considered a special case of stochastic interventions in which the probability of the outcome is equal to zero or one for each individual [11].

Epidemiologists have defined stochastic counterfactuals since the late 1980's [11–13], and more recently Muñoz and van der Laan proposed the stochastic intervention model (SIM) parameter, which extends PIM parameters to allow the intervention variable in the estimated counterfactuals to be assigned stochastically [14]. In contrast with PIM parameters, SIM parameters build in the random nature of exposure or uptake of an intervention. Thus, they are frequently a more realistic choice of parameter when evaluating a program or policy. Both PIM and SIM parameters can be defined for any type of intervention variable (e.g. binary, continuous) and are particularly useful with continuous intervention or exposure variables. In certain cases, SIM parameters are less susceptible to estimation-related problems than other commonly used approaches with continuous interventions, such as parametric regression models, or marginal structural models (MSMs) [15–17].

The purpose of this chapter is to illustrate how to estimate and interpret the population attributable fraction and its more modern variants (population intervention model and stochastic intervention model parameters) using simulated and empirical datasets. I will utilize a motivating example which applies these methods to understand the potential impact of a public health intervention that was imperfectly deployed. Specifically, using both the simulated and empirical datasets I will explore whether a program would have yielded better outcomes if it had been implemented perfectly. Frequently when interventions are deployed at a large scale or in a real world setting, their implementation is imperfect due to challenges in delivering interventions at scale or in obtaining high uptake among participants [18–23]. When implementation is imperfect, estimates of an intervention's effect will be closer to the null than would be expected if the intervention had been implemented perfectly. The parameters discussed in this chapter can be used to estimate the effect of the intervention if it had been implemented perfectly. Specifically, if information about the quality of program delivery is available, one can use these parameters to estimate if outcomes would have been better on average if all individuals received a better implemented intervention.

This chapter is organized as follows: I describe the research question and define parameters to estimate to answer this question. I then describe the simulation of several datasets that mimic data that would be produced by a sanitation program with imperfect implementation. The data simulated fall under four different scenarios which mimic the differing levels of data quality that may arise in empirical settings. Next, I describe the steps used to estimate and interpret each parameter from a single simulation under a scenario with ideal data. The description of parameter estimation is intended to be didactic, and accompanying code for these steps is provided in the Appendix. After describing the estimation of all four parameters in detail, I discuss the performance of three of the parameters across 1,000 simulations under more realistic data scenarios likely to arise in empirical settings, and I describe the variability of a single parameter estimate by examining its bootstrap distribution. I then estimate these parameters in an empirical dataset from an evaluation of a large-scale water, sanitation, and hygiene program implemented by UNICEF and the Government of Bangladesh in rural Bangladesh.

4.2 Data simulated

I have simulated data which mimics data from the empirical data from the UNICEF program described in forthcoming sections. I will use the simulated data to examine the effect of the implementation quality of a hypothetical sanitation program on a target outcome of interest. The variables used in the simulation are listed in Table 1. In this hypothetical program, participants received an improved latrine as well as promotion for use and maintenance of the latrine. Community health workers deliver the health promotion messages, and their interactions with participants can be of varying quality levels. Some community health workers may visit more frequently, and some may have better adherence to their job duties, and these differences result in a range of implementation quality. The key outcome of interest is "exclusive toilet use", which is commonly measured by asking respondents whether they have recently defected in bushes or fields rather than inside a latrine. By providing improved latrines, the program aims to eliminate the practice of open defecation (i.e., 100% exclusive toilet use).

The hypothetical study to answer this research question is cross-sectional. Table 4.1 lists the variables used in the simulation. I created a continuous index of implementation quality (A) in which a value of 0 indicates that the intervention was not implemented, a value of 50 indicates that the intervention was partially implemented (e.g. latrine installed but no promotion occurred), and a value of 100 indicates perfect implementation. The outcome of interest (Y) is exclusive toilet use. Regional poverty level (W) is a potential confounder. The research question is: What would the difference in the probability of open defection have been if all individuals had received a well-implemented program compared to the probability under the observed distribution of implementation quality?

I generated four simulated datasets with n=1,000 observations for each of the variables listed in Table 4.1. The dataset in Scenario 1 was designed to have ideal conditions, and the datasets in the remaining scenarios had features frequently found in observational data which could complicate parameter estimation. Table 4.2 lists the data generating distributions for the four scenarios.

Table 4.1: Variables simulated

Variable	Notation	Description
exclusive toilet use	Y	1=exclusive toilet use, 0=open defecation
Implementation quality	A	Continuous measure in which 0 indicates poor quality and 100 indicates good quality
Regional poverty level	W_1	1=live in an area in the 1st quartile of regional poverty
	W_2	1=live in an area in the 2nd quartile of regional poverty
	W_3	1=live in an area in the 3rd quartile of regional poverty
	W_4	1=live in an area in the 4th quartile of regional poverty

- In Scenario 1, the scenario with ideal conditions, Y is evenly distributed, A is evenly distributed, and strata of W are well-balanced with no sparse strata.
- In Scenario 2, Y is evenly distributed, A is right-skewed, and strata of W are well-balanced. This scenario could occur if the range of implementation quality observed was low on average and if no individual received a perfectly implemented program.
- In Scenario 3, Y is rare, A is normally distributed, and strata of W are well-balanced. This situation could occur if the association between implementation quality and exclusive toilet use is low and if open defecation is commonly practiced in the study population.
- In the Scenario 4, Y is evenly distributed, A is normally distributed, and strata of W are imbalanced. If the intervention was targeted to regions with a high poverty level, there would be fewer observations in the strata of W for lower poverty level regions. Furthermore, it is possible that certain values of A might never be observed within certain strata of W. For example, in the highest poverty region, it is possible that there would be no values of A above the median because implementation quality was poorer there.

In each scenario, I included interaction between A and W. For this hypothetical research question, it is plausible that the association between implementation quality and no open defection or other outcomes would be modified by regional poverty level. For example, in less impoverished areas, it is possible that the program would be better implemented because potential community health workers might be better educated or because transportation systems were better and thus community health workers can more easily reach intervention recipients.

Table 4.2: Summary statistics and data generating distributions

Scenario 1: Y and A are evenly distributed, W not sparse		
	n (%)	Data generating distribution
Y(%)	611 (61.1)	P(Y) $\sim$ Bin(n=1000, p=Logistic($-3 + 0.05 \times A + 0.005 \times W_2 + 0.007 \times W_3 + 0.009 \times W_4$ $+0.01 \times A \times W_2 + 0.03 \times A \times W_3 + 0.05 \times A \times W_4$))
Mean A (SD)	51.1 (13)	P(A) $\sim$ N($\mu = 50 + 0.5 \times W_2 + 1.5 \times W_3 + 2 \times W_4, \sigma = 13.3$)
W_1(%)	268 (26.8)	P(W_1)=0.25
W_2 (%)	248 (24.8)	P(W_2)=0.25
W_3 (%)	242 (24.2)	P(W_3)=0.25
W_4 (%)	242 (24.2)	P(W_4)=0.25
Scenario 2: Y is evenly distributed, A is right-skewed, W not sparse		
	n (%)	Data generating distribution
Y(%)	472 (47.2)	P(Y) $\sim$ Bin(n=1000, p=Logistic($-3 + 0.05 \times A + 1.7 \times W_2 + 1.9 \times W_3 + 2.3 \times W_4$ $+0.01 \times A \times W_2 + 0.03 \times A \times W_3 + 0.05 \times A \times W_4$))
Mean A (SD)	19.1 (19)	P(A) $\sim$ Beta($\alpha = 0.3 + 0.4 \times W_2 + 0.7 \times W_3 + 0.8 \times W_4, \beta = 3$) $\times$ 100
W_1(%)	268 (26.8)	P(W_1)=0.25
W_2 (%)	248 (24.8)	P(W_2)=0.25
W_3 (%)	242 (24.2)	P(W_3)=0.25
W_4 (%)	242 (24.2)	P(W_4)=0.25
Scenario 3: Y is rare, A is normally distributed, W not sparse		
	n (%)	Data generating distribution
Y(%)	85 (8.5)	P(Y) $\sim$ Bin(n=1000, p=Logistic($-7 + 0.05 \times A + 0.005 \times W_2 + 0.007 \times W_3 + 0.009 \times W_4$ $+0.01 \times A \times W_2 + 0.03 \times A \times W_3 + 0.05 \times A \times W_4$))
Mean A (SD)	51.1 (13)	P(A) $\sim$ N($\mu = 50 + 0.5 \times W_2 + 1.5 \times W_3 + 2 \times W_4, \sigma = 13.3$)
W_1(%)	268 (26.8)	P(W_1)=0.25
W_2 (%)	248 (24.8)	P(W_2)=0.25
W_3 (%)	242 (24.2)	P(W_3)=0.25
W_4 (%)	242 (24.2)	P(W_4)=0.25
Scenario 4: Y and A are evenly distributed, W sparse		
	n (%)	Data generating distribution
Y(%)	274 (27.4)	P(Y) $\sim$ Bin(n=1000, p=Logistic($-7 + 0.1 \times A + 0.5 \times W_2 + 0.0007 \times W_3 + 0.006 \times W_4$ $+0.01 \times A \times W_2 + 0.00003 \times A \times W_3 + 0.2 \times A \times W_4$))
Mean A (SD)	51.4 (13)	P(A) $\sim$ N($\mu = 51 + 0.5 \times W_2 + 1.5 \times W_3 + 2 \times W_4, \sigma = 13.3$)
W_1(%)	628 (62.8)	P(W_1)=0.6
W_2 (%)	288 (28.8)	P(W_2)=0.3
W_3 (%)	58 (5.8)	P(W_3)=0.07
W_4 (%)	26 (2.6)	P(W_4)=0.03

4.3 Parameters

I consider three parameters in this analysis: the population attributable fraction (PAF), population intervention model (PIM), and stochastic intervention model (SIM). Each of these parameters involves two scenarios: the observed scenario and a counterfactual scenario which improves upon the observed scenario. For this study, I have defined parameters on the additive scale (i.e. differences), however, it is also possible to define them on the relative scale (i.e., ratios). The definition of the parameters is very similar; what distinguishes them is the decision rule (d) used to assign the counterfactual level of the intervention (A).

1. The **population attributable fraction (PAF)** estimated in this simulation compares the observed expectation of exclusive toilet use (Y) to the expectation if everyone had received a perfectly implemented program (A=100):

$$\psi^{\mathrm{PAF}} = E[Y_{d^{\mathrm{PAF}}(a)}] - E[Y] \tag{4.1}$$

where $d^{\mathrm{PAF}} = a$. In this study, I consider a counterfactual in which if everyone had received a perfectly implemented program (a=100). This definition of the population attributable fraction draws upon the variant of the PAF defined by Levin, which was defined as $(E[Y] - E[Y_{A=0}])/E[Y]$ because the measure was applied to scenarios in which one imagines removing a deleterious exposure ($A = 0$) [4]. However, since I evaluated an intervention that is intended to be beneficial, I redefined the PAF to yield a positive result. To make the PAF comparable with the other parameters discussed below, I also did not divide by $E[Y]$. Although the parameter I have defined is not a fraction, I will refer to it as the PAF.

2. The **population intervention model (PIM) parameter** compares the observed expectation of the outcome (Y) to the expectation under an improved scenario.

$$\psi^{\mathrm{PIM}} = E[Y_{d^{\mathrm{PIM}}(a,A)}] - E[Y] \tag{4.2}$$

where d^{PIM} is a decision rule defined as follows:

$$d^{\mathrm{PIM}}(a, A) = A \cdot I(A \geq a) + a \cdot (1 - I(A \geq a))$$

where a is a pre-defined level in the data of relevance to the research question and $I(A \geq a)$ is an indicator of whether the observed value of A is greater than or equal to a for a given observation. In this study, I consider a counterfactual in which everyone with an observed value of implementation quality below the 75th percentile was reassigned to the value at the 75th percentile and everyone with an observed value above the 75th percentile retained their observed value ($A = a^{\mathrm{PIM}}$). For example, if the 75th percentile of the distribution of A is 60, then the decision rule is $d^{\mathrm{PIM}}(60, A) = A \cdot I(A \geq 60) + 60 \cdot (1 - I(A \geq 60))$. If an individual's implementation quality is observed to be

50, they are reassigned to 60; if their observed value is 80, their value remains 80. This parameter can accomodate control for potential confounders and averages over them in order to estimate the parameter for the whole study population.

3. The **stochastic intervention model (SIM) parameter** compares the observed expectation of the outcome (Y) to the expectation of Y under an improved scenario, but the counterfactual in the improved counterfactual is defined stochastically.

$$\psi^{\text{SIM}} = E[Y_{d^{\text{SIM}}(a,A)}] - E[Y] \tag{4.3}$$

where d^{SIM} is a decision rule defined as follows:

$$d^{\text{SIM}}(a, A) = A \cdot I(A \geq a) + A^* \cdot (1 - I(A \geq a))$$

$$A^* \sim P_n(A|A > a)$$

The random variable A^* is the value drawn from the empirical distribution of A above the pre-determined value a. In this study, the improved scenario of interest is one in which everyone with an observed value of implementation quality below the 75th percentile was reassigned to a value drawn from the empirical distribution of A above the 75th percentile and everyone with an observed value above the 75th percentile retained their observed value. For example, if the 75th percentile of the distribution of A is 60 and an individual's implementation quality is observed to be a score of 50, their value of A is drawn from the empirical distribution of A ranging from 60 to the maximum observed value of A. If their observed value is 80, their value of A remains 80.

4.4 Simulation procedures

First, I estimated each parameter of interest under each of the four data scenarios. To understand the performance of the parameters under the four scenarios of interest, I repeated the simulation 1,000 times for each parameter and each scenario. I calculated the true parameter estimates and estimated the proportion of times the 95% confidence interval for a given simulation's parameter estimate ($\hat{\psi}$) included the true value of each parameter (ψ) (i.e., the coverage probability). I calculated the true value of the PAF for each scenario analytically. To determine the true value of the PIM and SIM parameters, I estimated each in a simulation with 1,000,000 observations. I plotted the kernel density estimates of the 1,000 estimates of each parameter under each scenario using a Gaussian kernel. To understand the variability of parameter estimates, I bootstrapped a single parameter estimate with 1,000 replicates and plotted the kernel density estimates of the bootstrapped parameter estimates of each parameter under each scenario. I used **R** version 3.0.2.

4.5 Estimation

The target parameters in this study can be estimated using methods from the causal inference literature including G-computation [24], inverse probability of treatment weighting (IPTW), and targeted maximum likelihood, a form of double robust estimation [25]. Ahern et al. and Snowden et al. describe estimation of the average treatment effect using G-computation [26, 27], and Fleischer et al. estimated the population intervention model parameter using IPTW [28]. Muñoz and van der Laan utilized IPTW and TMLE in their paper proposing SIM parameters [14]. I have used G-computation to estimate the parameters in this chapter because it is straightforward and easy to implement relative to other estimators, and two didactic papers targeting epidemiologists already describe estimation in detail [26, 27]. For simplicity, I will use maximum likelihood to estimate these parameters, but it is important to note that these parameters can also be estimated with alternative semi-parametric techniques which make fewer assumptions about the data and may produce less biased and in some cases less variable results [25].

At least two didactic papers targeting epidemiologists have been published describing the steps used to implement G-computation. I provide a brief description of G-computation estimation steps here and refer readers to Ahern et al. and Snowden et al. for more details [26, 27]. G-computation separates estimation into two steps: 1) the estimation of the mean of Y ($E[Y|A,W]$) and 2) the estimation of the parameter using the estimate of $E[Y|A,W]$. G-computation can be envisioned as a process which imputes unobserved counterfactual outcomes for each unit. For example, to estimate the population attributable fraction with G-computation, one would impute the probability of exclusive toilet use for each individual if they had received a perfectly implemented program (A=100). These counterfactual outcomes are unobserved because in the observed data, many individuals did not have a value of A equal to 100. G-computation allows estimation of $E[Y|A,W]$ to be performed using methods other than regression, such semi-parametric, data-adaptive approaches [29]. Thus, the parameter is not necessarily tied to its estimation method and can be defined to estimate a quantity optimal to answer one's particular research question. For simplicity, I only estimated $E[Y|A,W]$ using maximum likelihood estimation, but semi-parametric estimation techniques could also be used which require fewer difficult-to-validate assumptions than logistic regression.

In this section I describe the estimation steps and interpret the results for each parameter under scenario 1, in which data is intended to be well-behaved. In later sections, I describe and interpret results for the remaining scenarios. Table 4.4 contains the parameter estimates for a single simulation under each scenario. In each scenario, I used the correct data distribution when specifying statistical models: I included main effects for A and W as well as an interaction between A and each level of W. However, it is important to note that in practice, one never knows the true data generating distribution, and thus parameter estimates will likely be at least somewhat biased.

Estimation steps

Step 1: Estimate the expectation of the outcome (Y) under the improved counterfactual scenario

1. **Impute values of A for each individual under the improved scenario**. Table 4.3 below contains hypothetical data in which the 75th percentile of the observed distribution of A is equal to 70, the observed value of A, and the imputed values (a^*) vary by parameter. For the PAF, all individuals' values are imputed to be 100. For the PIM parameter, the value of A for the first five individuals is below the 75th percentile, so their values are imputed as 70. The remaining individuals retain their original values of A. For the SIM parameter, the approach is similar as for PIM except that the first five individuals' values are imputed by drawing from the empirical distribution between 70 and 98.

Table 4.3: Hypothetical observed and counterfactual values of A

	Observed	Counterfactual a^*		
i	a	PAF	PIM	SIM
1	0	100	70	76
2	13	100	70	95
3	55	100	70	88
4	69	100	70	71
5	74	100	74	74
6	82	100	82	82
7	90	100	90	90
⋮	⋮	⋮	⋮	⋮
n	98	100	98	98

2. **Estimate the expectation of Y controlling for A and W ($\hat{E}[Y|A,W]$)**. I used a generalized linear model (GLM) with a binomial family and logit link (i.e., logistic regression) because the outcome is binary, however different models can be chosen depending on the nature of the data in the study. As a toy example, let us consider a model in which I assume there is a single binary confounder (W):

$$\ln\left(\frac{E[Y=1|A,W]}{1-E[Y=1|A,W]}\right) = \beta_0 + \beta_1 A + \beta_2 W$$

For example, the results of this step could be: β_0=-3, β_1=0.05, β_2=0.005.

3. **Estimate the probability of the outcome *for each individual* under the improved counterfactual scenario within strata of** W ($\hat{E}[Y_i|A_i, W_i]$) using the imputed values of A and the coefficients from the previous step. For example, for observation $i = 1$ in Table 4.3, if W=1, the predicted probability of exclusive toilet use for the PAF counterfactual equals $1/(1+\exp\{-(-3+0.05\times 100+0.005\times 1)\}) = 0.88$, the probability for the PIM counterfactual equals $1/(1+\exp\{-(-3+0.05\times 70+0.005\times 1)\}) = 0.62$, and for the SIM counterfactual it equals $1/(1 + \exp\{-(-3 + 0.05 \times 76 + 0.005 \times 1)\}) = 0.69$.

4. **Estimate the *average* probability of** Y **under the improved counterfactual scenario** within strata of W. This step allows us to estimate a parameter for the whole study population rather than within strata of W.

$$\hat{E}_W\left[\hat{E}[Y|A_{d(a,A)}, W]\right] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i|A_{d(a,A_i)}, W_i]$$

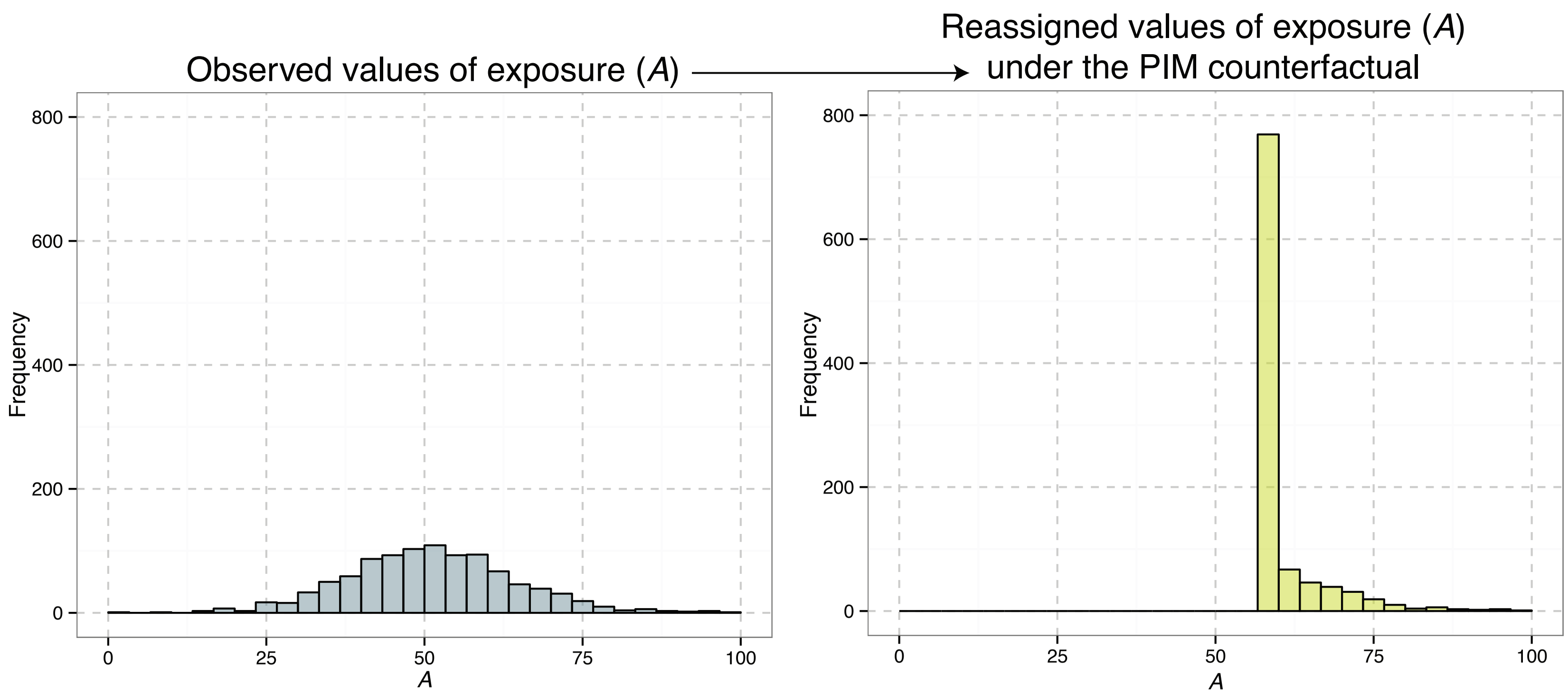
Step 2: Estimate the empirical mean of the outcome under the observed scenario ($\hat{E}[Y]$).

One can simply take the empirical mean of Y_i.

Step 3: Subtract the observed mean of the outcome (from Step 2) from the mean under the improved scenario (from Step 1).

$$\hat{\psi} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} E[Y_i|A_{d(a,A_i)}, W_i]}_{\text{Improved scenario}} - \underbrace{\frac{1}{n}\sum_{i=1}^{n} Y_i}_{\text{Observed scenario}}$$

Figure 4.1: Observed and counterfactual distributions of A under PIM - Scenario 1



In this scenario, *A* ranges from 0 to 100, and the 75th percentile is equal to an implementation quality score of 54.

Under the improved counterfactual scenario for the PIM parameter, all individuals with an observed value of *A* less than 54 were reassigned to *A*=54. Those with observed values of *A* above 54 retained their observed value. In this counterfactual distribution of *A*, approximately 750 observations now have a value of *A*=54, but the shape of distribution at values of *A* greater than 54 is identical to the shape of the observed distribution of *A*.

Observed *A*   Counterfactual *A* under PIM

Figure 4.2: Observed and counterfactual distributions of A under SIM - Scenario 1

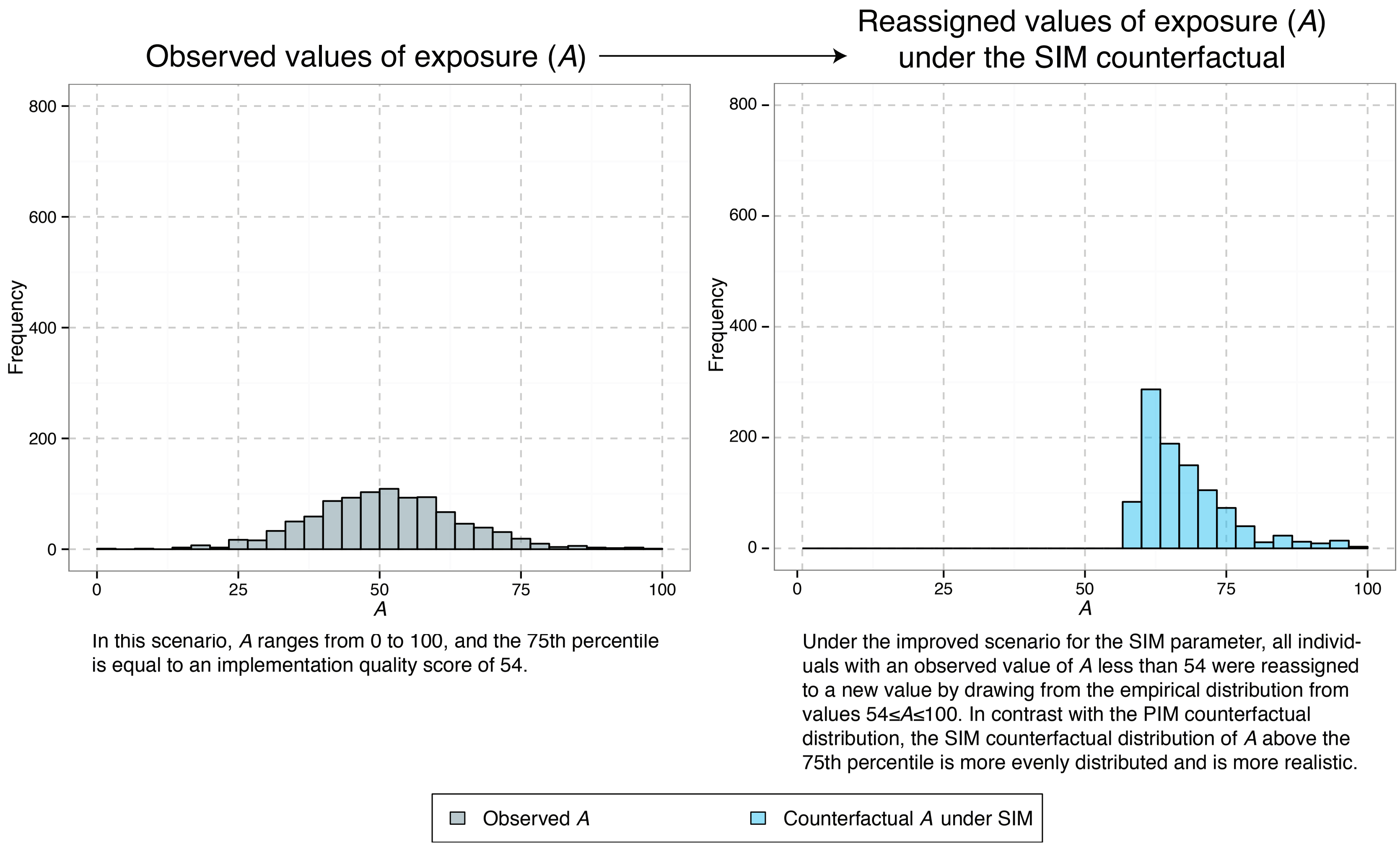
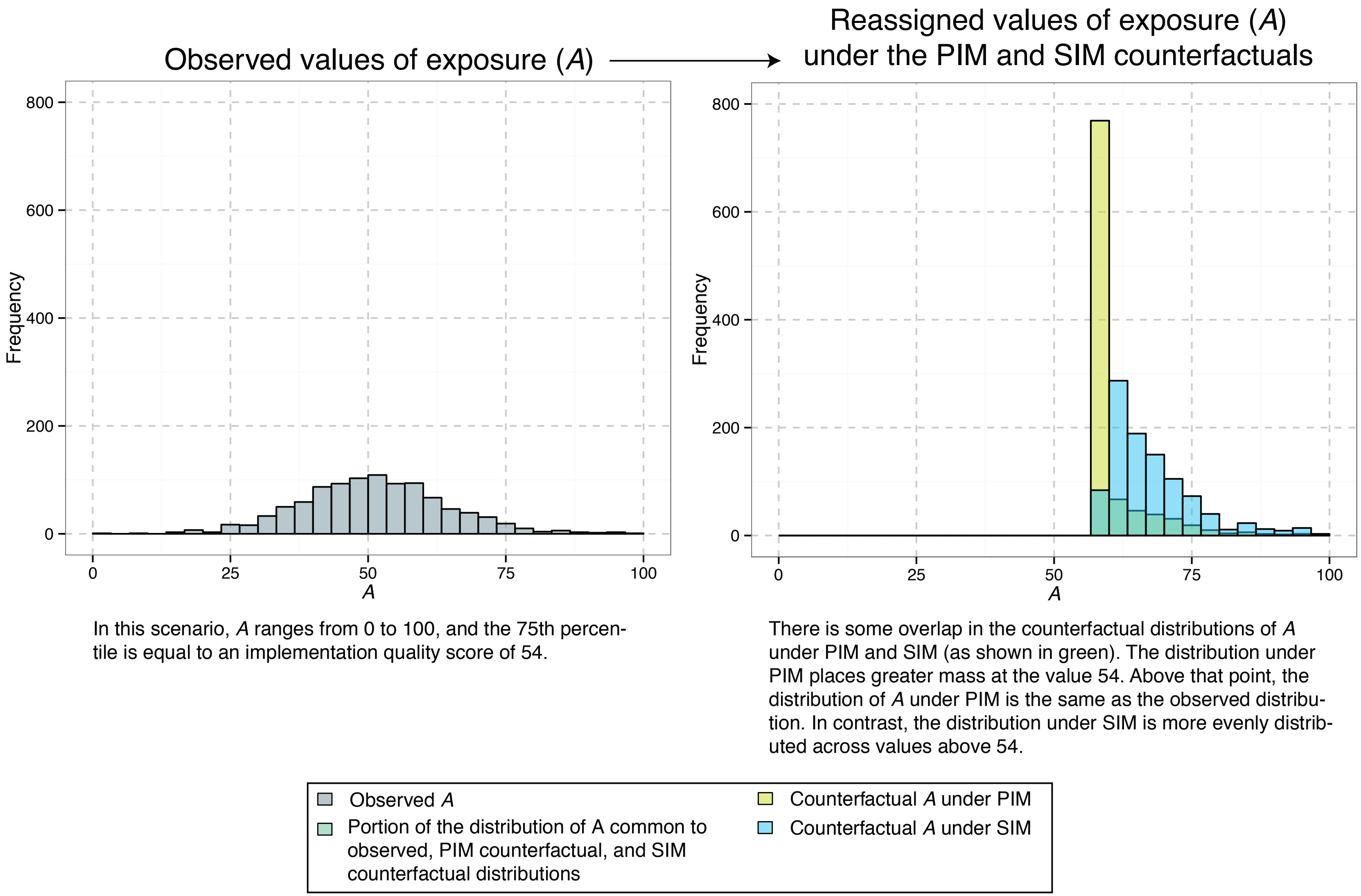


Figure 4.3: Observed and counterfactual distributions of A under PIM and SIM - Scenario 1



Interpretation of parameter estimates for Scenario 1

The data simulated for Scenario 1 were intended to be ideal, and parameter estimates are shown in the first row of Table 4.4. In this scenario, the estimate of the PAF was 0.303. Thus, if all individuals had received a perfectly implemented program, I would expect that the probability of exclusive toilet use would be 30.3% higher than it was observed to be. In comparison, the PIM parameter estimate for scenario 1 is 0.118. To interpret this parameter, if everyone had received a program with at least the 75th percentile of implementation quality, the probability of exclusive toilet use would have been approximately 11.8% higher than it would have been at the observed level of implementation quality. It is to be expected that the PIM parameter estimate is lower than the PAF estimate because the PIM parameter compares the observed scenario to an improved scenario, while the PAF compares the observed scenario to a scenario with 100% perfect implementation.

The SIM parameter estimate was 0.163: if everyone had received a program at least the 75th percentile of implementation quality, the probability of exclusive toilet use would have been approximately 16.3% higher than it would have been at the observed level of implementation quality. At first glance, the interpretation of PIM and SIM parameters are identical, but the underlying counterfactual data are generally more realistic for SIM because the distribution of A more closely resembles what one might see in practice. Generally, the SIM estimate will exceed the PIM estimate because there are more observations with higher values of A, whereas in the PIM counterfactual many values are assigned to the 75th percentile. The standard errors were the smallest for the PIM parameter followed by the SIM parameter. It is possible that the higher SE of the PAF reflects the greater extent of model extrapolation in this parameter estimate. The PIM is likely less variable than the SIM because the counterfactual distribution of A is less variable. In Figures 4.1, 4.2, 4.3 and C.1, which show the observed and counterfactual distributions of A under PIM and SIM.

Table 4.4: Results from a single simulation

	PAF	PIM	SIM
	$\hat{\psi}$ (SE)	$\hat{\psi}$ (SE)	$\hat{\psi}$ (SE)
Scenario 1: Y and A are evenly distributed, W not sparse	0.303 (0.028)	0.118 (0.013)	0.163 (0.017)
Scenario 2: Y is evenly distributed, A is right-skewed, W not sparse	0.496 (0.045)	0.161 (0.017)	0.272 (0.032)
Scenario 3: Y is rare, A is normally distributed, W not sparse	0.341 (0.055)	0.041 (0.006)	0.091 (0.014)
Scenario 4: Y and A are evenly distributed, W sparse	0.698 (0.018)	0.155 (0.011)	0.296 (0.021)

4.6 Comparison of parameter estimates across 1,000 simulations

In this section, I compare the results for the simulations repeated 1,000 times across scenarios 1 through 4, which are intended to illustrate more realistic data quality scenarios that may occur with observational data. Results from a single simulation are listed in Table 4.4, and Figure 4.4 displays the kernel density estimates for the parameter under estimates from 1,000 simulations under each scenario. Table 4.5 lists the coverage probability for each parameter and scenario, which is defined as the proportion of times the 95% confidence interval for each of the parameter estimates included the true parameter value. Ideally the coverage probability should equal the 95% for a 95% confidence interval in order to indicate that the type I error rate is properly accounted for; if it is less than 95%, it could indicate that a slightly greater than type I error rate and that one will incorrectly conclude there is a statistically significant finding more than 5% of the time.

Scenario 1: ideal data

In Scenario 1, in which Y was evenly distributed and A was normally distributed with no sparse covariate strata, the distribution of estimates for each parameter was approximately normal for each parameter. For each parameter, the distribution of estimates from the 1,000 simulations was centered on the true value, indicating minimal bias in the parameter estimates. The width of the distribution indicates the variability of parameter estimates. The distribution of PIM estimates was the narrowest, followed by the distribution for SIM and then the PAF. The PIM was likely less variable than the SIM parameter because the counterfactual distribution of A is less variable because many values were reassigned to the 75th percentile. The variability of the SIM and PAF estimates was similar. It is possible that this is because in the counterfactual distribution of A under PAF all values were set to 100, where there were few observed values in the data. Thus, PAF estimation required more model extrapolation than estimation of the PIM and SIM parameters. The coverage probabilities (Table 4.5) were similar for each parameter and were all close to 95%, indicating that the type I error rate over the 1,000 simulations was close to 5% as desired.

Scenario 2: A is right-skewed

In Scenario 2, the estimates of each parameter were on the whole evenly distributed, but they were slightly more variable but the distribution of SIM estimates was approximately normal. As shown in Figure C.1, the maximum value of A in Scenario 2 was 91, and many of the observed values of A equal 0. The distributions of each parameter were not centered around the true values, indicating that parameter estimates were more biased in this scenario. The distributions were also wider for each parameter than in Scenario 1, and they followed the same pattern (PIM was the narrowest followed by SIM and PAF). In this scenario the distribution of PAF estimates was somewhat wider than in Scenario 1. Scenario 2 required more model extrapolation than Scenario 1 because there were fewer observed values of A around the 75th percentile than in Scenario 1. There were approximately 50 observations within 2 units of the 75th percentile of A in Scenario 1 compared to approximately 20 observations in Scenario 2. Thus, there was less observed information to support the estimation of the probability of Y around this value of A in Scenario 2, which may explain the bias observed.

Scenario 3: Y is rare

In Scenario 3, as shown in Table 4.2, the prevalence of Y in this scenario was 8.5% compared to 61.1% in Scenario 1. The distributions of parameter estimates were centered around true values for each parameter, indicating little bias. The PIM was considerably less variable than in Scenario 1, and the SIM was slightly less variable, but PAF estimates were much more variable, as indicated by the wide distribution of PAF estimates. The PAF's high variability likely reflects the fact that in this scenario the probability of Y is very small at most values of A, but the probability increases rapidly when A exceeds the 75th percentile. Thus, when reassigning the value of A to 100 in the counterfactual scenario, the probability of Y will change greatly for observations starting with low values of A but will not change substantially for those with higher values of A. Thus, the PAF as I have defined it may be highly variable when the outcome is rare.

Scenario 4: W is sparse

In Scenario 4, W was sparse within strata of Y and A. Specifically, there were very few observed values of W_3 and W_4. For each parameter, the distributions were centered around the true values, indicating little bias. The variability of the parameter estimates was also quite similar to Scenario 1. These results suggest that having sparse covariate strata, at least as observed in this simulation, does not strongly impact the bias or variance of estimates of the PAF, PIM, or SIM.

To further evaluate the variability of parameter estimates, I plotted the distribution of bootstrapped parameter estimates from a single simulation for each parameter and scenario (Figure C.2). The patterns are similar to those observed in Figure 4.4. With the exception of the distribution of PIM in Scenario 1, the distributions were smooth and unimodal, indicating that the assumptions needed to use bootstrap were likely met in this analysis.

4.7 Application to empirical data: evaluation of SHEWA-B program

In this section I illustrate the application of these parameters to an empirical dataset from an evaluation of a water, sanitation, and hygiene program in rural Bangladesh. From 2007

Figure 4.4: Distribution of parameter estimates over 1,000 repetitions for each scenario

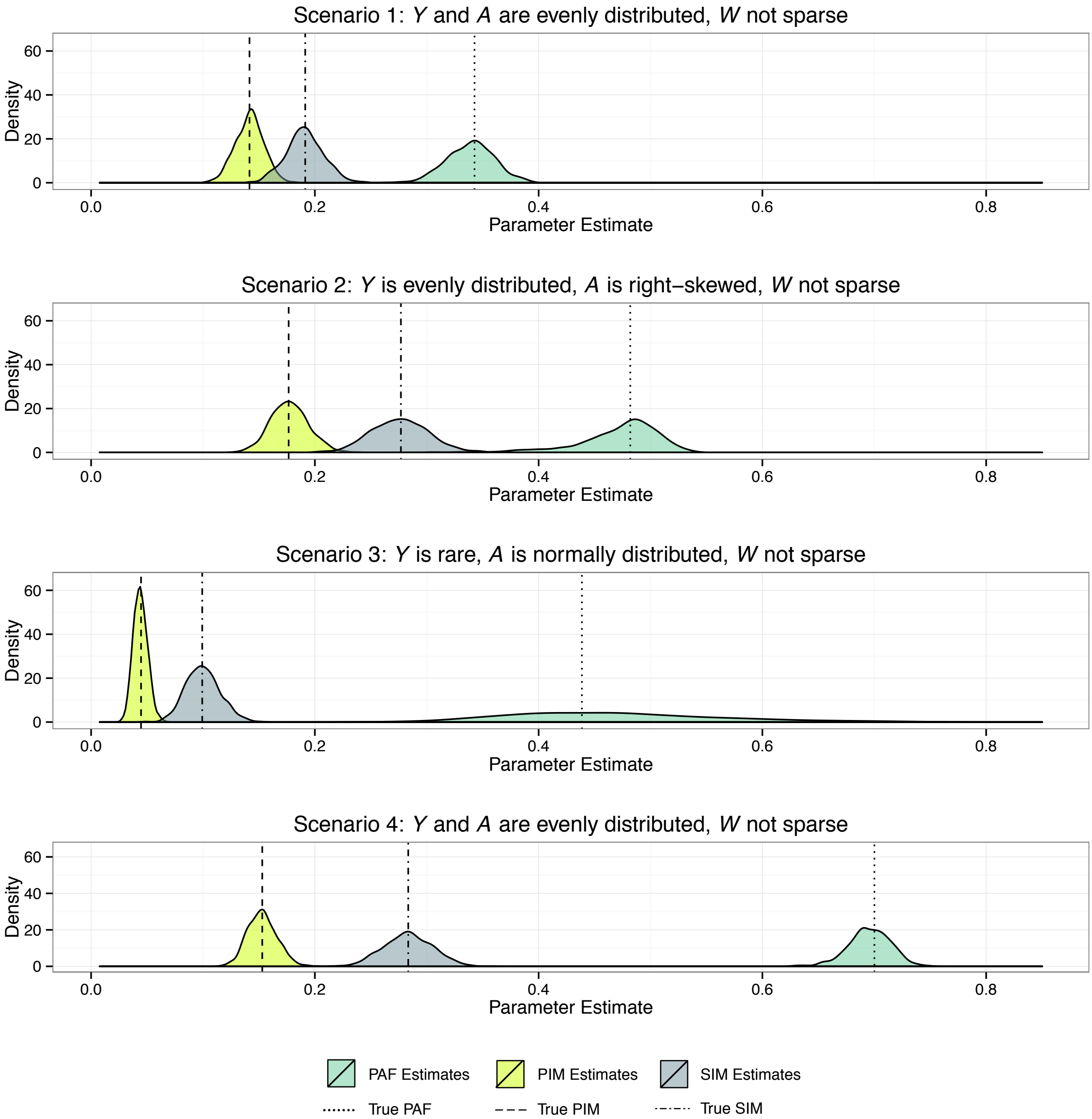


Table 4.5: Coverage probability* for each parameter and scenario

	PAF	PIM	SIM
Scenario 1: Y and A are evenly distributed, W not sparse	95%	96%	96%
Scenario 2: Y is evenly distributed, A is right skewed, W not sparse	95%	94%	94%
Scenario 3: Y is rare, A is normally distributed, W not sparse	89%	93%	93%
Scenario 4: Y and A are evenly distributed, W not sparse	97%	93%	93%

*The coverage probability is the proportion of times the 95% confidence interval $\hat{\psi}_i \pm z_{1-\alpha/2} \times SE(\hat{\psi}_i)$ includes the true parameter value (ψ) for $i = 1, ..., 1000$ bootstrap replicates.

to 2012 UNICEF and the Government of Bangladesh delivered a program called Sanitation Hygiene Education and Water Supply in Bangladesh (SHEWA-B), which targeted approximately 20 million beneficiaries in rural Bangladesh. The intervention promoted safe hygiene and sanitation practices and its objective was to reduce diarrhea and other water and hygiene-related diseases among the poorest in rural Bangladesh. In 2009, an interim assessment conducted by the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) found little improvement in measures of hygiene sanitation behavior and no improvement in child diarrhea and respiratory illness [30]. To explore whether these results reflected suboptimal intervention design versus suboptimal intervention implementation, I collaborated with ICDDR,B to develop an index of SHEWA-B implementation quality and estimated the extent to which target outcomes would have improved if SHEWA-B had been better implemented. In this section I estimate the PAF, PIM, and SIM parameters as defined above with the SHEWA-B evaluation data. Details on the data collected, implementation quality index, and evaluation design are described in Chapter 1. To briefly summarize, the index of implementation quality ranged from 0 to 100 in possible values, and it was calculated at the village cluster level since the intervention was essentially delivered at the village level. I also consider exclusive toilet use as the outcome in this analysis.

Figure 4.5 shows the observed distribution of implementation quality in SHEWA-B. The distribution resembles that in Scenario 2 of the simulation: there are many clusters with a index score of zero (i.e., the intervention was not implemented), and none had a score of 100 (perfect implementation). Table 4.6 lists the estimates and standard errors for each parameter in this empirical dataset. The PAF parameter compares the probability of exclusive toilet use in the observed data to the probability if everyone received perfect implementation quality (index=100). The PAF estimate was 0.034, which indicates that the probability of exclusive toilet use would have increased by 3.4% if implementation was perfect. The PIM and SIM compare the observed scenario to one in which all clusters received implementation quality equal to the 75th percentile of the index in this dataset (index=42). The PIM estimate was 0.008 and the SIM estimate was 0.013. As in the simulation, the PIM estimate is the closest to the null, and the PAF has the largest effect size. All estimates are quite small and suggest that even if implementation quality had been better, the program would have not have greatly increased exclusive toilet use.

I have estimated the PAF with a counterfactual assuming perfect implementation to be consistent with the definition of the PAF in the simulation. However, because there were no village clusters with an observed implementation quality index value at that level, it is generally not advisable to do so because it requires one to extrapolates beyond the observed data used to generate the model. It is also very unlikely that a program targeting such a large population would achieve perfect implementation in all areas, so the counterfactual for improved in the PIM and SIM parameters is more realistic than the counterfactual of perfect implementation in the PAF. In this case, the PIM and SIM estimates both have small effect sizes, and their standard errors are identical to the third decimal place. The SIM is slightly more realistic than the PIM; one would expect that the SHEWA-B program, which mostly promoted health behaviors, would result in a probability of increased exclusive toilet use rather than deterministically resulting in exclusive toilet use. Thus, given the identical variability of the SIM and PIM estimates and the more realistic definition of the SIM counterfactual, the SIM parameter is preferable in this setting.

Figure 4.5: Distribution of implementation quality in empirical example

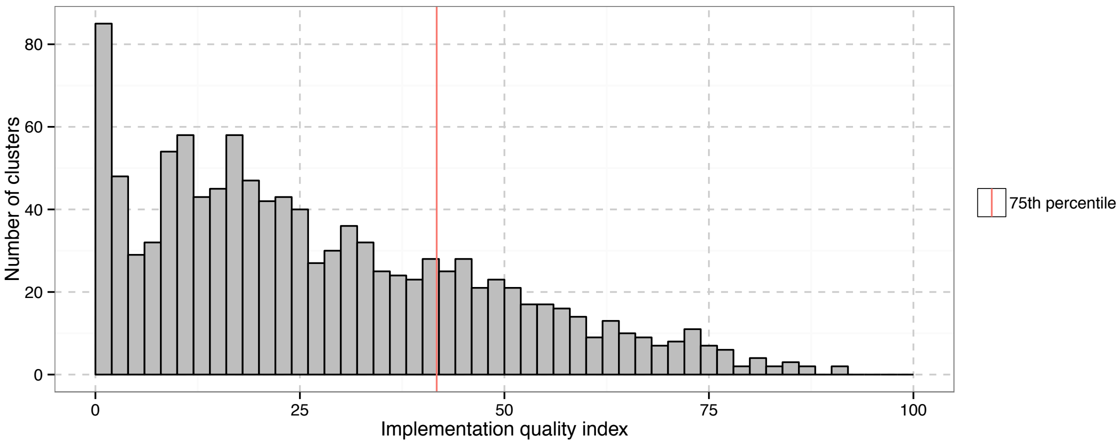


Table 4.6: Empirical results

Parameter	Point estimate	SE
PAF	0.034	0.011
PIM	0.008	0.002
SIM	0.013	0.002

4.8 Discussion

Key findings

In this study, I explored methods of estimating the population-level effect of a public health intervention using observational simulated and empirical data. I estimated the association between the quality of implementation of a hypothetical sanitation program and the probability of exclusive toilet use using four different parameters. The goal was to understand what the difference in the probability of the outcome would be under an improved scenario. Each of the three parameters answers the research question in a slightly different way – the population attributable fraction estimates how different the probability would have been if everyone received a perfectly implemented program. The PIM and SIM parameters estimate how different the probability would have been if more people received a better implemented program. The interpretation of PIM and SIM is quite similar, but the counterfactual distribution of SIM is more realistic because it accounts for the fact that individuals' responses to this behavior change intervention will follow some probability distribution rather than deterministically resulting in a particular response.

Repeating the simulation 1,000 times allowed of investigation of the properties of these parameters under different data conditions, some of which are likely to arise in observational data. I found that the PIM estimates were closer to the null than SIM parameter estimates in all scenarios. This is because the counterfactual distribution of the intervention variable had greater mass closer to the mean under the PIM counterfactual than under SIM counterfactual. The PIM estimates were less variable than SIM estimates across all scenarios. The PAF estimates' variability was similar to the SIM parameter's when under ideal data conditions and when the covariates were sparse, but when the intervention variable was skewed or the outcome was rare, the PAF was highly variable. All parameters were biased when the distribution of the intervention variable was skewed, and PAF was particularly susceptible to bias in that scenario. Thus, investigators interested in applying these parameters in their data should carefully explore features of their data prior to parameter definition and estimation, and if the distribution of the intervention variable is skewed, estimate these parameters with caution.

In studies evaluating the population-level effects of health interventions, randomized trials are typically considered the optimal study design. However, they are often subject to a number of limitations, such as limited generalizability and measurement of a limited set of outcomes, and they are not feasible for a number of important research questions. While observational designs are subject to a number of limitations as well, thoughtful application of methods such as those I have demonstrated here can yield highly generalizable findings of great utility to public health practitioners and policymakers. Common critique of analyses of observational data are that the inference relies upon a statistical model rather than upon the study design [31] and that the choice of which quantity to estimate is driven by the statistical model used rather than by the research question [25]. This is more often the case for observational studies than trials because many trials can be analyzed non-parametrically if their randomization was effective, whereas observational designs typically employ statistical models to control for potential confounders. While statistical approaches can never remedy a poorly designed study, this study demonstrates an approach to analyzing observational data that can overcome such pitfalls through careful definition of parameters that are tailored to the specific research question and evaluation of the extent to which the parameter can be estimated with one's data.

Estimation methods for different study designs

The parameters and estimation methods used here can be applied to any observational study design, although estimation techniques vary by design. The key quantities needed to estimate the PAF, PIM, or SIM are the expectation of the outcome ($E[Y]$) and the expectation of the outcome conditional on the exposure or intervention ($E[Y|A,W]$). Here I have considered a cross-sectional study design in which sampling was independent of exposure and outcome status. Thus, both $E[Y]$ and $E[Y|A,W]$ can be estimated directly from the data using maximum likelihood or other semi-parametric approaches for independent and identically distributed data. In cohort studies, if the follow-up time is equal for all subjects and the sampling is independent of exposure, these parameters can be estimated as if the study was cross-sectional. However, if the probability of exposure is set by design, neither quantity can be directly estimated without corrections for the sampling probabilities. If follow-up time varies, $E[Y|A,W]$ must be estimated using survival analysis techniques, such as Poisson regression or Cox models [6, 7]. In case-control studies, neither $E[Y]$ or $E[Y|A,W]$ can be estimated directly from the data because sampling is conditional on outcome status. An alternative formulation of the PAF is recommended, and it can be used to estimate the PIM and SIM parameters as well: $(P(A)(RR-1))/(1+P(A)(RR-1))$, where $P(A)$ is the probability of the intervention and RR is the measure of association (e.g. relative risk, odds ratio). $P(A)$ can typically be estimated from the controls [5–7]. Estimation of the RR requires the analytical technique appropriate to the particular type of case-control design (e.g. case-cohort, nested case-control, etc.) [7]. Estimation of these parameters is also possible in trials, but for the reasons discussed above, the interpretation of the parameter will depend on the design and features of each particular trial (e.g. nature of intervention allocation, compliance, etc.).

Application of stochastic intervention models to other research questions

As discussed above, stochastic intervention parameters are frequently more realistic than deterministic ones. Stochastic intervention model parameters are broadly applicable to other research questions, and Table 4.7 contains a list of such questions. These may include interventions which target biologic measures, such as CD-4 T-cell counts, environmental measures, such as exposure to diesel exhaust, behavioral variables, such as hours of physical activity per week, and measures of intervention coverage, such as the proportion of the population with access to piped water and sewerage. Another useful application of these models is to estimate the potential impact of a health intervention for planning purposes. For example, prior to scaling up an intervention, one could estimate the extent to which the prevalence of disease would decrease if the intervention were fully scaled up. This kind of analysis lends itself well to analysis of many publicly available datasets, such as Demographic and Health Surveys datasets.

Table 4.7: Other potential research questions for which stochastic intervention model parameters may be appropriate

Topic	Question
HIV	How would patient outcomes vary under treatment regimes initiated based on differing thresholds of CD-4 T-cell counts? [10]
Recreational water	Would the risk of gastrointestinal illness be significantly lower than its current level if mean concentration of fecal indicator bacteria was always below the EPA recommended level?
Nutrition	How would the prevalence of coronary heart disease differ if the whole population consumed trans fatty acid consumption at the level equal to or less than the lowest quintile of consumption compared to the current level?
Occupational health	How would the risk of lung cancer differ among workers if the number of years working in jobs with high exposure to asbestos was reduced to 10 years compared to the observed number of years? [11]
Global burden of disease	How much lower would the mortality rate attributable to water and sanitation be if everyone in the world had access to fully functional sewerage and piped water?
Chronic disease	How would the prevalence of Type II diabetes differ if the whole population increased the amount they exercised per week from their current level by one hour?
Neglected tropical diseases	If 100% of the population received mass drug administration for five years, would the prevalence of lymphatic filariasis decrease sufficiently from its current level to interrupt transmission?

SIM parameters can be defined in a variety of ways and estimated with data from a range of study designs. In this simulation, I have defined a SIM parameter for a continuous intervention variable, but they can also be defined for binary or categorical intervention variables. I have illustrated the estimation of these parameters with observational data, but they can also be estimated with data from other study designs (e.g. case-control, randomized controlled trial). However, one of the advantages of analyses with observational data over data from a trial is that the investigator can define counterfactuals of interest, whereas in trials counterfactuals are fixed by design [16].

Estimation techniques

I have used G-computation for the estimation of the parameters, but other estimation techniques can be used, such as inverse probability of treatment weighting (IPTW) and targeted maximum likelihood (TMLE). Fleischer et al. used IPTW to estimate PIM parameters [28], and Muñoz and van der Laan used both IPTW and TMLE to estimate SIM parameters [14]. All of these methods decouple estimation of the mean adjusting for potential confounders and estimation of the parameter itself. While I have used maximum likelihood estimation in this simulation, this separation of estimation steps allows for estimation of the mean of the outcome using other semi-parametric techniques which make fewer assumptions about the underlying data, such as SuperLearner [29] or other machine learning algorithms. I conducted my analysis in **R**, but for investigators interested in using Stata, Stata version 13 now includes commands for inverse probability weighting and doubly robust methods, and similar packages are available in **R** as well.

Assumptions underlying estimation

If one aims to make causal inferences, a set of assumptions are required. I refer readers to Ahern et al. (2009) for a discussion of these assumptions in the context of G-computation [26]. One of these assumptions – the positivity assumption – is important to consider even when the goal is not causal inference because it affects both the bias and variance of estimates [15]. The assumption states that there is a positive probability of each level of the intervention variable within each level of covariate strata. When the exposure or intervention variable is continuous, technically the experimental treatment assignment assumption is always violated because one cannot observe all levels of treatment in each strata of the covariates. Thus, it is necessary to extrapolate beyond the observed data using a statistical model.

Fortunately, the positivity assumption can be assessed using the observed data by examining the distribution of A within strata of W. Figure C.3 shows the distributions of propensity scores (the probability of the intervention controlling for covariates) in each scenario. When the propensity score values are close to zero, there is a high probability of a violation of the positivity assumption [15]. In Figure C.3, very few observations have a propensity score of zero in Scenarios 1, 3, and 4, but there are numerous observations with propensity scores close to zero in Scenario 2. Thus, there is likely an ETA violation in Scenario 2, in which the distribution of the intervention variable was skewed. This is not surprising given that the range of the intervention variable in this scenario did not reach the maximum possible value, and the distribution was right-skewed. When there is a positivity violation, to estimate the probability of the outcome at certain values of the intervention variable requires reliance upon a statistical model and extrapolation beyond the data in the model. Doing so warrants caution since it is possible that the model parameters would have differed if more data was available.

Petersen et al. describe methods of diagnosing and responding to positivity violations in detail [15]. In the case of SIM and PIM parameters as I have defined them, if one observes a skewed distribution of the intervention variable, one potential solution is to define the parameter such that the counterfactual level above which the intervention variable is reassigned is in a region where there is a reasonable amount of support in the data. For instance, in Scenario 2, I could define the SIM and PIM counterfactuals around the 60th percentile instead of the 75th percentile of the observed distribution of A in an attempt to avoid reliance on extrapolation of model results generated from a small number of observations at higher values of A. While redefining the parameter in this way can help avoid severe positivity violations, they also change the interpretation of the parameter – using the 60th instead of the 75th percentile would mean comparing the observed scenario to a less improved counterfactual scenario.

Another assumption discussed at length in the causal inference literature is consistency, which refers to whether or not assignment to a particular intervention level or value will always yield the same outcome [32–35]. This assumption is problematic when an intervention or exposure can result in a range of different responses. For example, there are a number of ways in which someone could have received poor implementation quality that may lead to differing probabilities of the outcome. If someone was disinterested in the program, their demeanor may have reduced the community health worker's motivation to try to change their behavior. Alternatively, someone may have been very interested in the health promotion offered by the community health worker but lived in a remote area that was difficult for the health worker to reach. The probability of exclusive toilet use is likely to be different for these two scenarios even under the same value of implementation quality. As with other assumptions made in causal inference, it is rarely the case that we can safely assume consistency, even when one is able to implement or manipulate the intervention of interest.

However, defining stochastic parameters can allow one to relax the consistency assumption, which can have advantages in a number of settings. This is because when counterfactuals are defined stochastically, it is no longer necessary for the potential outcome to be identical for a given level of treatment, but instead, the distribution of potential outcomes must be the same for a given level of treatment [33]. Comparing distributions of potential outcomes allows for a range of outcomes to be observed, potentially caused by different mechanisms of obtaining a given level of A. The consistency assumption also arises in mediation analyses. In order to decompose the total effect of an intervention into direct and indirect effects, the consistency assumption is needed [36]. Stochastic counterfactuals can also relax this assumption when assessing mediation.

Conclusion

In conclusion, there are a number of modern parameters akin to the population attributable fraction that may be appropriate to estimate for a wide range of research questions. These parameters are particularly useful in evaluating potential effects of exposure or intervention that are measured as continuous variables. For any study design or research question, it is important to thoughtfully define one's parameter of interest and carefully assess whether it can be estimated with one's observed data. While observational data are never perfect, when collected and analyzed thoughtfully, they can yield powerful findings that are more broadly applicable than results of a randomized trial.

Bibliography

[1] Duflo E, Glennerster R, and Kremer M. "Using randomization in development economics research: a toolkit". *Handbook of Development Economics.* Vol. 4. Elsevier, B.V., 2008. Chap. 61.

[2] Farmer P. *Clinical trials and global health equity.* http://globalhealth.thelancet.com/2013/07/08/clinical-trials-and-global-health-equity. 2013.

[3] Rothman KJ, Greenland S, and Lash TL. *Modern epidemiology.* Lippincott Williams & Wilkins, 2008.

[4] Levin M. The occurrence of lung cancer in man. *Acta-Unio Internationalis Contra Cancrum* 9:3 (1953), p. 531.

[5] Bruzzi P, Green SB, Byar DP, Brinton LA, and Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *American Journal of Epidemiology* 122:5 (1985), pp. 904–914.

[6] Greenland S and Drescher K. Maximum Likelihood Estimation of the Attributable Fraction from Logistic Models. *Biometrics* 49 (1993), pp. 865–872.

[7] Benichou J. A review of adjusted estimators of attributable risk. *Statistical Methods in Medical Research* 10:3 (2001), pp. 195–216.

[8] Morgenstern H and Bursic ES. A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. *Journal of Community Health* 7:4 (1982), pp. 292–309.

[9] Hubbard A and van der Laan M. Population Intervention Models in Causal Inference. *Biometrika* 95 (2008), pp. 35–47.

[10] Cain L, Robins J, Lanoy E, Logan R, Costagliola D, and Hernán M. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *The International Journal of Biostatistics* 6:2 (2010).

[11] Robins J and Greenland S. The probability of causation under a stochastic model for individual risk. *Biometrics* 45 (1989), pp. 1125–1138.

[12] Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology* 125:5 (1993), pp. 761–768.

[13] Robins JM and Greenland S. Comment on “Causal Inferences Without Counterfactuals”. *Journal of the American Statistical Association* 95:450 (2000), pp. 431–435.

[14] Muñoz ID and van der Laan M. Population intervention causal effects based on stochastic interventions. *Biometrics* 68:2 (2012), pp. 541–549.

[15] Petersen M, Porter K, Gruber S, Wang Y, and van der Laan M. Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research* 21:1 (2010), pp. 31–54.

[16] Robins JM, Hernan MA, and Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11:5 (2000), pp. 550–560.

[17] Diaz I and van der Laan M. Assessing the causal effects of policies: an example using stochastic interventions. *International Journal of Biostatistics* 9:2 (2013), pp. 161–174.

[18] Chopra M and Ford N. Scaling up health promotion interventions in the era of HIV/AIDS: challenges for a rights based approach. *Health Promotion International* 20:4 (2005), pp. 383–390.

[19] Durlak JA. Why program implementation is important. *Journal of Prevention & Intervention in the community* 17:2 (1998), pp. 5–18.

[20] Durlak JA and DuPre EP. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology* 41:3-4 (2008), pp. 327–350.

[21] Mangham LJ and Hanson K. Scaling up in international health: what are the key issues? *Health Policy and Planning* 25:2 (2010), pp. 85–96.

[22] Rowe AK, Savigny D de, Lanata CF, and Victora CG. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *The Lancet* 366:9490 (2005), pp. 1026–1035.

[23] Subramanian S, Naimoli J, Matsubayashi T, and Peters DH. Do we have the right models for scaling up health services to achieve the Millennium Development Goals? *BMC Health Services Research* 11:1 (2011), p. 336.

[24] Robins J. A new approach to causal inference in mortality studies with a sustained exposure period: application to control of the healthy worker survivor effect. *Mathematical Modeling* 7:9 (1986), pp. 1393–1512.

[25] van der Laan MJ and Rose S. *Targeted learning: causal inference for observational and experimental data*. Springer, 2011.

[26] Ahern J, Hubbard A, and Galea S. Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. *American Journal of Epidemiology* 169:9 (2009), pp. 1140–1147.

[27] Snowden JM, Rose S, and Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *American Journal of Epidemiology* (2011).

[28] Fleischer N, Fernald L, and Hubbard A. Estimating the potential impacts of intervention from observational data: methods for estimating causal attributable risk in a cross-sectional analysis of depressive symptoms in Latin America. *J Epidemiol Community Health* 64 (2010), pp. 16–21.

[29] van der Laan MJ, Polley EC, et al. Super learner. *Statistical Applications in Genetics and Molecular Biology* 6:1 (2007), pp. 1–23.

[30] Huda TMN, Unicomb L, Johnston RB, Halder AK, Yushuf Sharker MA, and Luby SP. Interim evaluation of a large scale sanitation, hygiene and water improvement programme on childhood diarrhea and respiratory disease in rural Bangladesh. *Social Science & Medicine* 75:4 (2012), pp. 604–611.

[31] Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine* 26:1 (2007), pp. 20–36.

[32] Cole S and Frangakis C. The Consistency Statement in Causal Inference: A Definition or an Assumption? *Epidemiology* 20:1 (2009), pp. 3–5.

[33] VanderWeele T. Concerning the Consistency Assumption in Causal Inference. *Epidemology* 20:6 (2009), pp. 880–883.

[34] Pearl J. On the Consistency Rule in Causal Inference: Axiom, Definition, Assumption, or Theorem? *Epidemiology* 21:6 (2010), pp. 872–875.

[35] Petersen M. Compound Treatments, Transportability, and the Structural Causal Model: The Power and Simplicity of Causal Graphs. *Epidemiology* 22:3 (2011), pp. 378–381.

[36] VanderWeele T and Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface* 2 (2009), pp. 457–468.

Chapter 5

Conclusion

5.1 Key Findings

Findings of Chapter 2: The interaction of deworming, improved sanitation, and household flooring and soil-transmitted helminth infection in rural Bangladesh

Objective 1: to estimate the prevalence of STH infection among children and women of childbearing age in rural Bangladesh

No systematic surveys have been done to estimate STH prevalence in rural Bangladesh since national school-based MDA was initiated in 2008. I found that 40% of school-aged children, 26% of pre-school aged children, and 30% of women of childbearing age in rural Bangladesh in late 2012 had an STH infection. There were very few moderate or heavy STH infections in the study population. Because intensity of infection is the major driver transmission, the low intensity observed may indicate that transmission is waning in this study population, potentially as a result of the MDA program. The WHO recommends estimating prevalence after 5 years of school-based MDA [1]. For populations receiving MDA in which prevalence is between 10 and 50%, WHO recommends continuing MDA at its previous frequency for four more years and to "reinforce measures for safe water, sanitation and health education" [1]. Thus, my findings suggest that the MDA program has been successful in reducing STH prevalence in rural Bangladesh, continuation of the MDA program in conjunction with improvement of WASH conditions is necessary to further reduce STH prevalence.

Objective 2: to estimate associations with deworming, hygienic latrines, and finished floors and STH infection

I found protective associations with deworming, hygienic latrine access, and finished floors for *Ascaris* and hookworm prevalence, but associations were close to null for *Trichuris* prevalence. The majority of the associations with hookworm prevalence were not statistically significant, but this could reflect the small sample size. The deworming results are consistent with those of randomized controlled trials [2–4]. The results for hygienic latrine access were not statistically significant for any organism and differ from those reported in the literature, but most studies used different definitions of latrine access/quality which are less likely to indicate whether latrines prevent fecal contamination of the environment [5]. For finished floors, the results are consistent with those in the literature [6–8].

Objective 3: to explore potential interactions between these exposures

For each type of helminth, the joint exposures were consistently associated with a lower prevalence than individual exposures. Unfortunately my estimates of the RERI and RPR were in most cases underpowered. Despite that, the consistent pattern of prevalence I found supports further exploration of improved sanitation and finished floors as complementary interventions to reduce STH transmission.

Objective 4: to estimate associations between cluster-level exposures and cluster-level STH prevalence

Although mathematical models would suggest that increasing cluster-level coverage of deworming, hygienic latrines, and finished floors would reduce cluster-level STH prevalence, I found no meaningful decrease in prevalence for any cluster-level exposure. These findings warrant further investigation. Population density may be an important effect modifier of these associations that should be accounted for in future analyses.

Findings from Chapter 3: Assessment of a national-scale water, sanitation and hygiene intervention in rural Bangladesh: Measuring the effect of implementation quality

Ojective: estimate the extent to which hygiene behavior and conditions may have improved if SHEWA-B had been better implemented

To pursue this objective, my team and I conducted an assessment of one of the largest WASH programs ever conducted. I found that implementation of SHEWA-B was suboptimal: the majority of respondents did not recall ever meeting a community health promoter (CHP). Low exposure to CHPs was the main factor resulting in suboptimal implementation quality. However, outcomes were only marginally better among households who had met a CHP. While the qualitative assessment found that some CHPs performed well in certain areas, the observed distribution of implementation quality suggests that implementation quality did not meet UNICEF's ideal in any area. Even though some outcomes were near UNICEF's targets, the modest associations between implementation quality and targeted knowledge and behavior outcomes suggest that these outcomes may be better attributed to factors other than participation in SHEWA-B. These findings highlight the difficulty of maintaining high quality implementation at scale.

Findings from Chapter 4: Advances in the estimation of the population attributable fraction: application of a causal inference technique to simulated and empirical datasets

Objective: illustrate how to estimate and interpret the population attributable fraction, population intervention model, and stochastic intervention model parameters using simulated and empirical datasets

For the simulation, I generated four datasets with varying levels of data quality. I focused on a hypothetical research question inspired by the question explored in Chapter 3, and I estimated the three parameters with each simulated dataset as well as the empirical dataset from Chapter 3. I also repeated the simulations 1,000 times to investigate the properties of each parameter. I provided a didactic description of how to estimate each parameter and discussed their differing interpretations and applicability to a range of research questions. Stochastic intervention model parameters account for the random nature of exposure or uptake of an intervention. As a result, they are often a more realistic choice of parameter when evaluating a program or policy.

5.2 Discussion

This dissertation describes methods and applications of such methods to the evaluation of population-level effects of water, sanitation, and hygiene interventions. The specific interventions considered include deworming, hygienic latrines, finished floors, and promotion of hygiene behaviors (e.g. handwashing). All interventions considered here have been found to be efficacious at a small-scale, and many are now being scaled up. There are several lessons from this dissertation which can be considered in planning future evaluations of population-level effects of large scale interventions:

1. **It is important to clearly define the intervention prior to implementation.** While in some cases it is very easy to do so (e.g. provision of a deworming tablet), for more complex interventions, such as behavior change programs, there are numerous ways in which an intervention could be defined. Definition in advance of evaluation is critical to ensuring that both a process and impact evaluation can be done thoughtfully.

2. **Measurement of implementation quality can shed light on reasons for sub-optimal estimates of intervention impact.** As discussed in Chapter 3, when a program is found not to have an impact, one often wonders whether the intervention itself could have been better designed or if the design was sufficient but implemen-

tation was flawed. To explore the latter question, measurement of implementation quality during the implementation process is critical.

3. **Assessment of interaction between interventions can identify strategies for reducing disease prevalence more sustainably.** As discussed in Chapter 2, interventions that affect the same disease transmission pathways (e.g. fecal-oral) are often considered separately because of silos within the field of public health practice and research (e.g. the WASH and neglected tropical disease sectors). Assessing whether interventions delivered together can yield greater reductions than the total reductions from either intervention alone can help determine whether interventions could be designed to be complementary. In many cases, complementary interventions could yield great savings in cost and resources and could more sustainably reduce disease burden.

4. **Careful selection of one's parameter of interest can strengthen inference from observational data.** For reasons discussed in previous chapters, observational designs are often most appropriate for evaluation of large-scale interventions. Even though analyses with observational data require control for potential confounders, one need not select the target parameter based solely on the statistical model used. Chapter 4 discussed three specific parameters relevant to understanding population-level effects of interventions and illustrated how to define and estimate parameters tailored to specific research questions.

As mentioned above, there has been increased funding and motivation for the scale-up of efficacious public health interventions since the establishment of the Millennium Development Goals (MDGs) in 2000 [9]. We are nearing 2015, the date the UN set to evaluate progress towards the goals defined in 2000. In 2010, the UN General Assembly held a meeting to discuss the post-2015 of called for both the scale-up and integration of efforts proven to be successful to improve maternal and child health, and it specifically called for the scale up of WASH interventions to reduce child mortality [16]. The analytic methods illustrated in this dissertation can be applied to the evaluation of many other public health interventions; such an application would help to generate evidence both to support prioritization of scarce resources and to measure progress toward public health goals (such as the Millennium Development Goals for 2015 as well as country-specific health targets). These methods are also highly applicable to many specific areas of public health. A few relevant examples include: evaluations of interventions targeted at HIV; evaluations of progress toward elimination of neglected tropical diseases; evaluations of population risks from heart disease or diabetes; and evaluations directed toward understanding the social underpinnings of health. Finally, these methods have direct applicability for estimation of the global burden of numerous diseases and the contributions of specific exposures to those diseases.

Bibliography

[1] World Health Organization. *Helminth control in school-age children: A guide for managers of control programmes*. 2011. URL: http://whqlibdoc.who.int/publications/2011/9789241548267_eng.pdf.

[2] Albonico M, Smith PG, Hall A, Chwaya HM, Alawi KS, and Savioli L. A randomized controlled trial comparing mebendazole and albendazole against Ascaris, Trichuris and hookworm infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 88:5 (1994), pp. 585–589.

[3] Vercruysse J, Behnke JM, Albonico M, Ame SM, Angebault C, Bethony JM, Engels D, Guillard B, Hoa NTV, Kang G, et al. Assessment of the anthelmintic efficacy of albendazole in school children in seven countries where soil-transmitted helminths are endemic. *PLoS Neglected Tropical Diseases* 5:3 (2011), e948.

[4] Steinmann P, Utzinger J, Du ZW, Jiang JY, Chen JX, Hattendorf J, Zhou H, and Zhou XN. Efficacy of single-dose and triple-dose albendazole and mebendazole against soil-transmitted helminths and Taenia spp.: a randomized controlled trial. *PLoS One* 6:9 (2011), e25003.

[5] Ziegelbauer K, Speich B, Mäusezahl D, Bos R, Keiser J, and Utzinger J. Effect of sanitation on soil-transmitted helminth infection: systematic review and meta-analysis. *PLoS Medicine* 9:1 (2012), e1001162.

[6] Faust EC and Giraldo LE. Parasitological surveys in Cali, Departamento del Valle, Colombia VI. Strongyloidiasis in Barrio Siloé, Cali, Colombia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 54:6 (1960), pp. 556–563.

[7] Hall A, Conway DJ, Anwar KS, and Rahman ML. Strongyloides stercoralis in an urban slum community in Bangladesh: factors independently associated with infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 88:5 (1994), pp. 527–530.

[8] Walker M, Hall A, and Basáñez MG. Individual predisposition, household clustering and risk factors for human infection with Ascaris lumbricoides: new epidemiological insights. *PLoS Neglected Tropical Diseases* 5:4 (2011), e1047.

[9] Sachs JD. Health in the developing world: achieving the Millennium Development Goals. *Bulletin of the World Health Organization* 82:12 (2004), pp. 947–949.

[10] Bloom G and Ainsworth P. *Beyond scaling up: pathways to universal access to health services*. STEPS Working Paper 40. STEPS Centre, 2010.

[11] Tangcharoensathien V and Patcharanarumol W. Commentary: Global health initiatives: opportunities or challenges? *Health Policy and Planning* 25:2 (2010), pp. 101–103.

[12] Lu C, Schneider MT, Gubbins P, Leach-Kemon K, Jamison D, and Murray CJ. Public financing of health in developing countries: a cross-national systematic analysis. *The Lancet* 375:9723 (2010), pp. 1375–1387.

[13] *World Development Report 1993: Investing in Health*. World Bank and Oxford University Press, 1993.

[14] *Scaling Up: From Vision to Large-Scale Change*. Management Systems International, 2012. URL: `http://www.msiworldwide.com/wp-content/uploads/Scaling-Up-Framework.pdf`.

[15] Duflo E. *Scaling Up and Evaluation*. 2004. URL: `http://economics.mit.edu/files/766`.

[16] 65/1 GA resolution. *Keeping the promise: united to achieve the Millennium Development Goals*. 19 October 2010. URL: `http://www.un.org/en/mdg/summit2010/pdf/outcome_documentN1051260.pdf`.

Appendix A

Appendix to Chapter 1

Figure A.1: Directed acyclic graph used to identify potential confounders

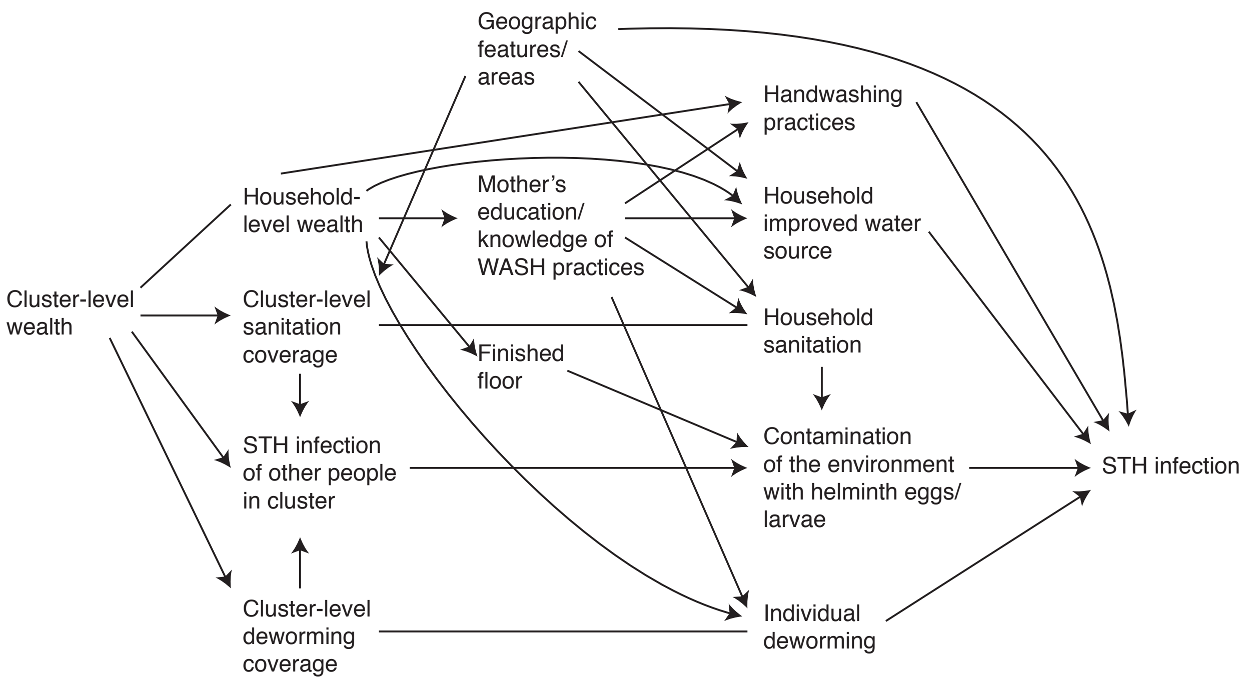


Table A.1: Prevalence ratios for improved vs. hygienic latrine access*

	Ascaris PR (95% CI)	Hookworm PR (95% CI)	*Trichuris* PR (95% CI)
Unadjusted prevalence ratios			
JMP improved sanitation	0.99 (0.76,1.30)	1.41 (0.92,2.14)	1.04 (0.84,1.29)
ICDDR,B hygienic latrine	0.78 (0.58,1.06)	0.60 (0.36,0.98)	0.93 (0.74,1.17)
Adjusted prevalence ratios†			
JMP improved sanitation	1.07 (0.81,1.41)	1.66 (1.08,2.54)	1.09 (0.88,1.36)
ICDDR,B hygienic latrine	0.88 (0.64,1.22)	0.71 (0.42,1.19)	1.03 (0.81,1.32)

* ICDDR,B developed a definition of "hygienic" latrines which differs from the WHO Joint Monitoring Programme (JMP) but may be a more accurate categorization of latrines that isolate feces from the environment for the types of sanitation found in Bangladesh. Hygienic latrines include flush latrines connected to piped sewer system, to septic tank, or off-set pit latrine, pit latrine with slab and functional water seal, pit latrine with slab, lid and no water seal, or a composting latrine. Unhygienic latrines are those that fail to effectively separate feces from the environment: flush latrine connected to canal or ditch, pit latrine without slab, pit latrine with slab, no or broken water seal or a hanging latrine.

This definition differs from the JMP definition in two ways. Hygienic latrines require a water seal or a lid on a pit to effectively separate collected faeces from the environment and does not consider sharing status of a latrine. "No access to a latrine" included households who reported no facilities, defecating in open spaces, fields or near water bodies. Field workers also recorded self reports of the latrine ownership and sharing status from the respondents.

† PRs estimated using log binomial regression and adjusted for age, sex, sub-district, household wealth, cluster-level wealth, and mother's education level

Table A.2: Means of each variable used in the principal components analysis by quintile of the index

	Quintile of wealth index				
Variable	1	2	3	4	5
Electricity	0.040	0.076	0.115	0.170	0.200
Almirah	0.015	0.045	0.112	0.176	0.198
Table	0.033	0.122	0.163	0.190	0.198
Chair	0.044	0.134	0.175	0.190	0.201
Clock	0.008	0.031	0.063	0.091	0.168
Khat	0.039	0.077	0.135	0.186	0.199
Chouki	0.141	0.144	0.130	0.130	0.137
Radio	0.001	0.001	0.004	0.008	0.025
Black and white TV	0.001	0.005	0.018	0.022	0.053
Color TV	0.000	0.007	0.015	0.062	0.137
Refrigerator	0.000	0.000	0.002	0.014	0.057
Bicycle	0.011	0.034	0.053	0.048	0.097
Motorcycle	0.000	0.000	0.001	0.003	0.045
Sewing	0.004	0.005	0.016	0.018	0.056
Mobile phone	0.083	0.138	0.170	0.194	0.201
Sofa	0.000	0.000	0.001	0.005	0.040
Car	0.000	0.001	0.002	0.004	0.010
Land	0.016	0.016	0.009	0.011	0.005
Homestead	0.178	0.189	0.190	0.198	0.199

Table A.3: Interaction between sanitation and deworming

	No deworming		Deworming				
	n/N	PR* (95% CI)	n/N	PR* (95% CI)	PR for deworming*	RERI† (95% CI)	RPR‡ (95% CI)
Ascaris							
No hygienic latrine	106/652	1.0	44/305	0.63 (0.46,0.87)	0.63 (0.46,0.87)		
Hygienic latrine	49/445	0.97 (0.69,1.37)	14/219	0.44 (0.25,0.77)	0.48 (0.26,0.88)		
PR for sanitation*		0.97 (0.69,1.37)		0.66 (0.35,1.24)			
						-0.43 (-2.40,0.48)	0.77 (0.35,1.53)
Hookworm							
No hygienic latrine	47/652	1.0	23/445	0.82 (0.50,1.32)	0.82 (0.50,1.32)		
Hygienic latrine	11/305	0.55 (0.28,1.07)	9/219	0.81 (0.38,1.73)	1.25 (0.50,3.12)		
PR for sanitation*		0.55 (0.28,1.07)		1.32 (0.58,3.00)			
						0.45 (-0.88,1.15)	1.53 (0.54,4.33)
Trichuris							
No hygienic latrine	130/652	1.0	107/445	1.20 (0.95,1.51)	1.20 (0.95,1.51)		
Hygienic latrine	70/305	1.20 (0.91,1.58)	35/219	0.87 (0.61,1.24)	0.70 (0.48,1.02)		
PR for sanitation*		1.20 (0.91,1.58)		0.73 (0.51,1.06)			
						-0.65 (-1.41,-0.13)	0.59 (0.35,0.92)

*PRs adjusted for age, sex, sub-district, household wealth, cluster-level wealth, and mother's education level

†Relative excess risk due to interaction (RERI). A RERI=0 indicates no interaction on the additive scale, RERI>0 indicates synergistic interaction on the additive scale for monotonic aPRs, RERI>1 indicating synergistic interaction on the additive scale for non-monotonic aPRs, and RERI<0 indicates antagonistic interaction on the additive scale.

‡Ratio of prevalence ratios

Table A.4: Interaction between finished floors and deworming

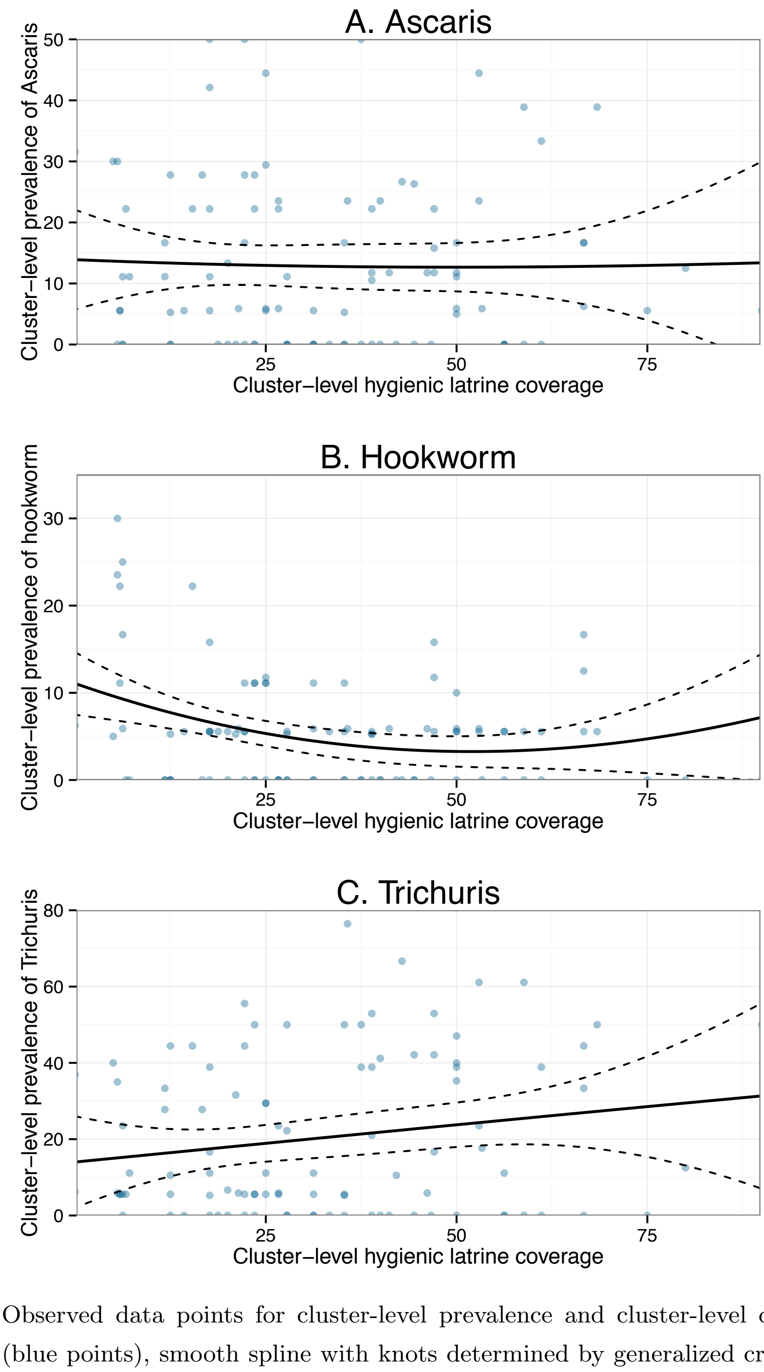
	No deworming		Deworming				
	n/N	PR∗ (95% CI)	n/N	PR∗ (95% CI)	PR for deworming∗	RERI† (95% CI)	RPR‡ (95% CI)
Ascaris							
Unfinished floor	141/847	1.0	59/572	0.60 (0.45,0.80)	0.60 (0.45,0.80)		
Finished floor	9/110	0.54 (0.28,1.03)	4/93	0.29 (0.11,0.79)	0.68 (0.24,1.91)		
PR for finished floors∗		0.54 (0.28,1.03)		0.51 (0.18,1.42)			
						0.51 (-3.65,2.00)	1.13 (0.25,3.67)
Hookworm							
Unfinished floor	55/847	1.0	31/572	0.94 (0.61,1.44)	0.94 (0.61,1.44)		
Finished floor	3/110	0.50 (0.15,1.70)	1/93	0.26 (0.03,1.98)	0.85 (0.07,9.94)		
PR for finished floors∗		0.50 (0.15,1.70)		0.37 (0.04,3.09)			
						-1.04 (–,–)	0.91 (–,–)
Trichuris							
Unfinished floor	177/847	1.0	126/572	1.05 (0.85,1.29)	1.05 (0.85,1.29)		
Finished floor	23/110	1.10 (0.72,1.68)	16/93	0.96 (0.59,1.54)	0.86 (0.49,1.53)		
PR for finished floors∗		1.10 (0.72,1.68)		0.94 (0.57,1.54)			
						-0.23 (-1.50,0.37)	0.83 (0.41,1.55)

∗PRs adjusted for age, sex, sub-district, household wealth, and cluster-level wealth

†Relative excess risk due to interaction (RERI). A RERI=0 indicates no interaction on the additive scale, RERI>0 indicates synergistic interaction on the additive scale for monotonic aPRs, RERI>1 indicating synergistic interaction on the additive scale for non-monotonic aPRs, and RERI<0 indicates antagonistic interaction on the additive scale.

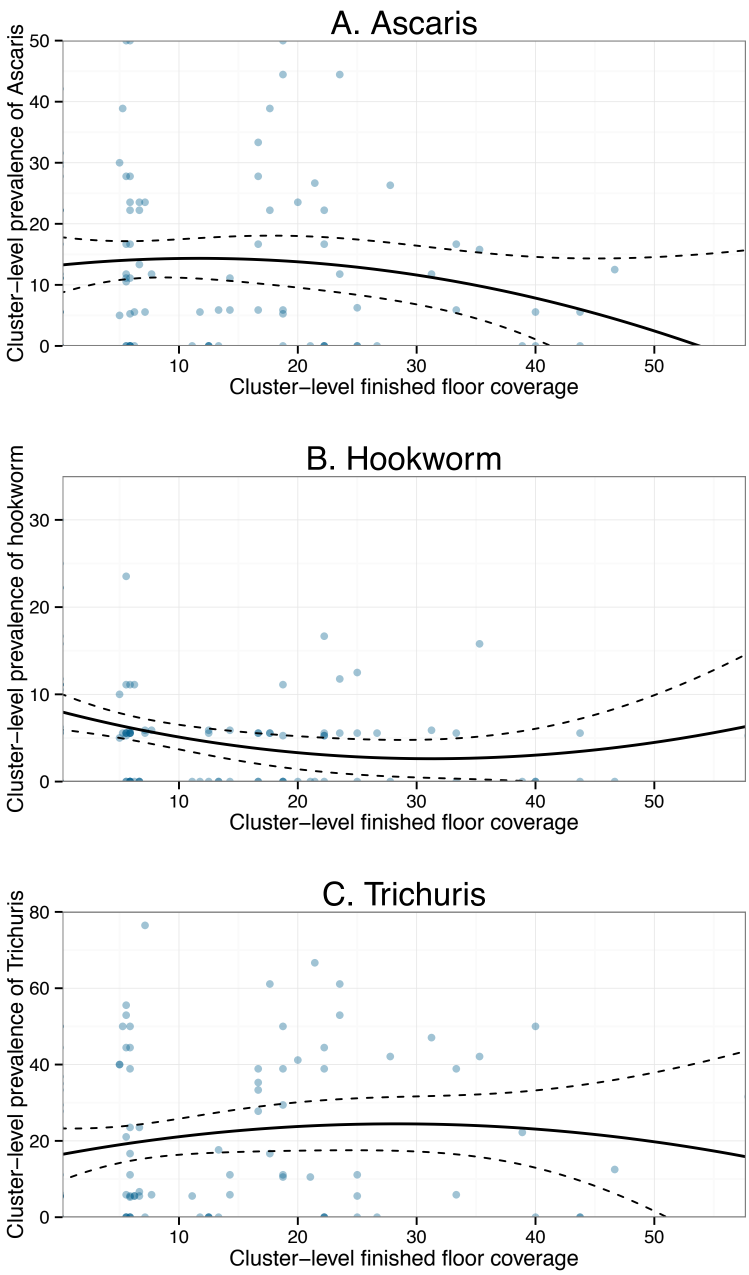
‡Ratio of prevalence ratios

Figure A.2: Cluster-level STH prevalence by cluster-level sanitation coverage



Observed data points for cluster-level prevalence and cluster-level deworming coverage (blue points), smooth spline with knots determined by generalized cross-validation (solid line), and point-wise 95% confidence bands (dashed lines). In B) the degrees of freedom were manually set to 7 because generalized cross-validation yielded a df greater than 15.

Figure A.3: Cluster-level STH prevalence by cluster-level finished floor coverage



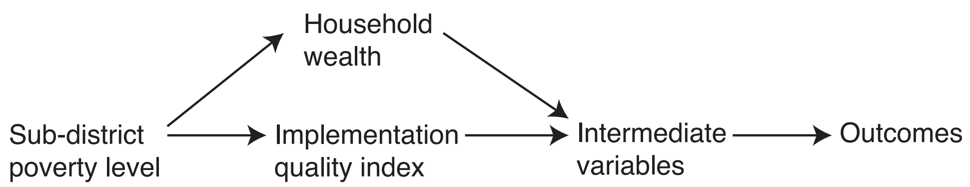
Observed data points for cluster-level prevalence and cluster-level finished floor coverage (blue points), smooth spline with knots determined by generalized cross-validation (solid line), and point-wise 95% confidence bands (dashed lines).

Appendix B

Appendix to Chapter 2

Figure B.1: Directed acyclic graphs used to identify potential confounders and select covariates for statistical models

A. Graph for caregiver knowledge of hygiene, sanitation and safe water messages, improved latrine access



B. Graph for households with latrines with no feces on latrine slabs or floors; open defecation; hygienic drinking water points; demonstrated handwashing behavior; observed hand cleanliness; household keeps drinking water in a covered container; households having soap and water at convenient handwashing place after defecation

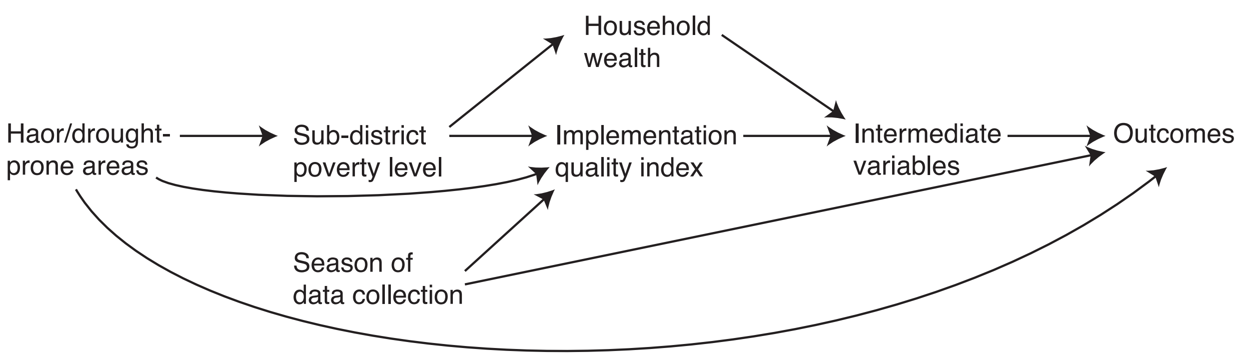


Table B.1: Planned and actual sample size

	Planned	Actual*	Response rate (%)
Cross-sectional survey			
Upazilas	58	58	
Clusters	1182	1182	
Households	33096	33027	99.7
Community hygiene promoter survey			
Upazilas	58	58	
Clusters	1182	1126	
CHPs	1164	1110	95.2

∗ In the first two sub-districts, we collected data in 31 clusters each rather than 20 clusters each.

Table B.2: Outcomes stratified by frequency of CHP visit

	At least one Never met a CHP	CHP visit in CHP visit	CHP visit in last 4 months	CHP visit in last month
Outcome	% (95%CI)	% (95%CI)	% (95%CI)	% (95%CI)
Private, improved latrine available	23 (21,24)	24 (22,25)	22 (20,24)	23 (21,25)
No feces on latrine slab or floor	51 (49,52)	50 (48,52)	48 (46,51)	51 (49,53)
No open defecation	93 (92,94)	94 (94,95)	94 (93,95)	95 (94,96)
Has dedicated handwashing location	55 (54,57)	62 (60,64)	60 (57,62)	57 (55,59)
Correct caregiver handwashing	57 (55,58)	64 (62,66)	64 (62,66)	67 (65,69)
Caregiver hands observed to be clean	44 (43,45)	42 (40,44)	43 (41,46)	44 (42,46)
Child hands observed to be clean	29 (28,30)	28 (27,29)	28 (26,30)	31 (29,33)
Received at least 1 W,S,H message from CHP	0 (–,–)	74 (72,76)	74 (72,76)	79 (78,81)
Respondent heard any messages from a SHEWA-B CHP	0 (–,–)	94 (93,95)	98 (98,99)	99 (99,99)
Drinking water container covered	42 (40,43)	43 (40,47)	43 (39,46)	49 (46,52)
Drinking water point is sanitary	30 (29,31)	26 (24,28)	26 (24,28)	26 (24,28)

Table B.3: Difference in probability of outcomes under observed implementation quality and under counterfactual scenario

Outcome	Effect Modifier*	n	Point Estimate	95% CI
Private, improved latrine available	None	1,126	0.9	(-0.0,1.8)
No open defecation	0-24% poverty	1,126	0.0	(-0.4,0.4)
	25-40% poverty	1,126	1.6	(0.9,2.5)
	31-36% poverty	1,126	0.0	(-0.7,0.9)
	37-55% poverty	1,126	1.1	(-0.0,2.2)
	Cool season	1,126	1.1	(0.6,1.7)
	Hot season	1,126	0.4	(-1.1,1.8)
	Rainy season	1,126	-1.9	(-3.9,-0.0)
No feces on latrine slab or floor	None	1,125	1.3	(0.2,2.3)
Clean caregiver hands	Cool season	1,126	-0.7	(-1.6,0.3)
	Hot season	1,126	-1.6	(-5.6,2.2)
	Rainy season	1,126	4.2	(1.3,6.6)
Clean child hands	None	1,126	0.7	(-0.1,1.6)
Dedicated handwashing location	None	1,126	1.3	(0.2,2.4)
Correct caregiver handwashing	0-24% poverty	1,126	0.5	(-2.1,2.9)
	25-40% poverty	1,126	4.1	(2.1,6.6)
	31-36% poverty	1,126	3.8	(2.4,5.5)
	37-55% poverty	1,126	5.9	(4.4,7.4)
Drinking water point is sanitary	None	1,126	-0.8	(-1.7,0.2)
Drinking water container covered	None	1,064	0.9	(-0.8,2.4)

* Effect modification was considered to be statistically significant when a model had at least two 95% confidence intervals for risk differences within strata of effect modification that do not overlap in models with statistical interactions. Main effects are shown for models for which effect modification was not statistically significant.

Appendix C

Appendix to Chapter 3

Figure C.1: Observed and counterfactual distributions of A under PIM and SIM - Scenario 2

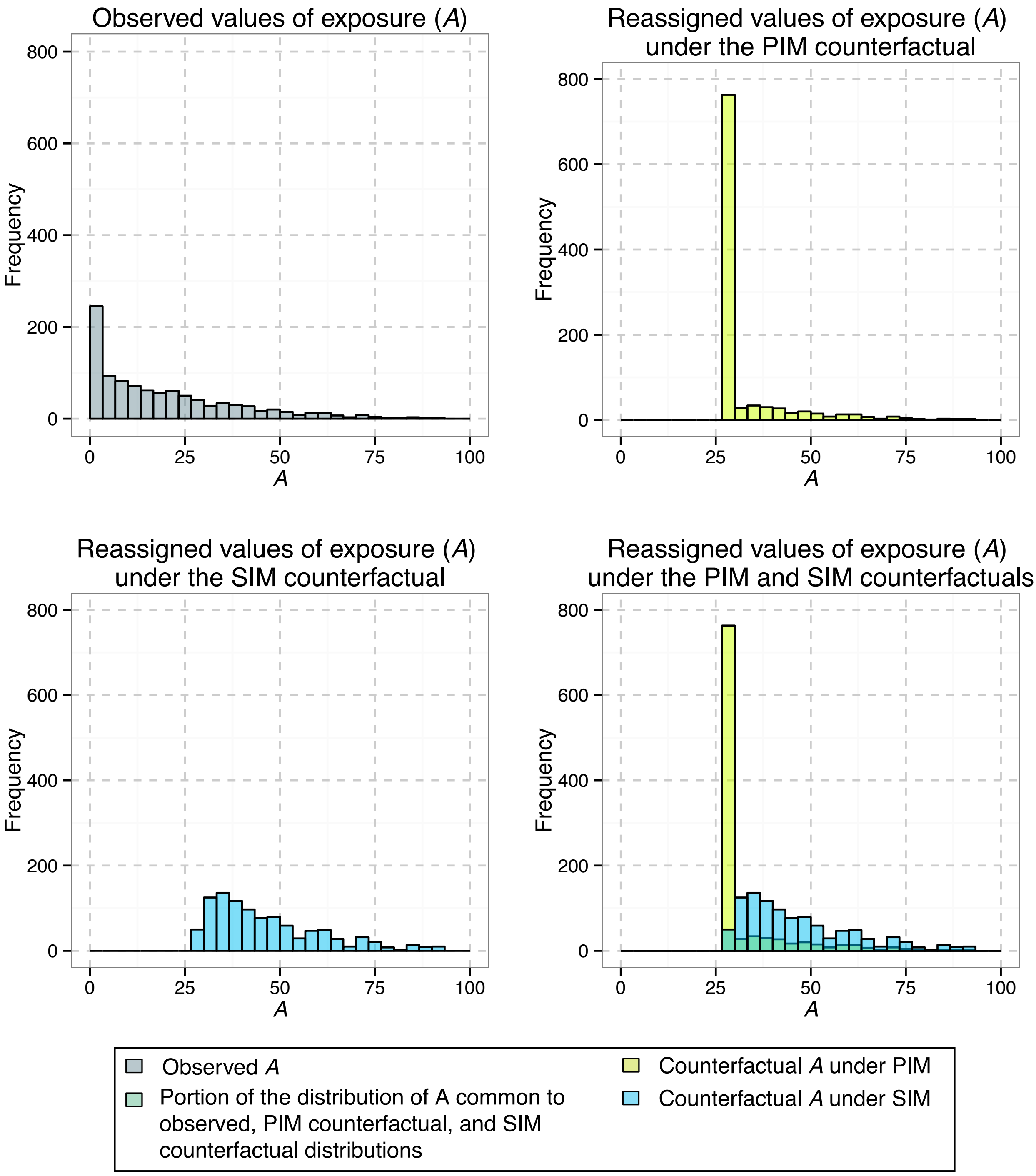


Figure C.2: Distribution of bootstrapped parameter estimates with 1,000 repetitions for each scenario

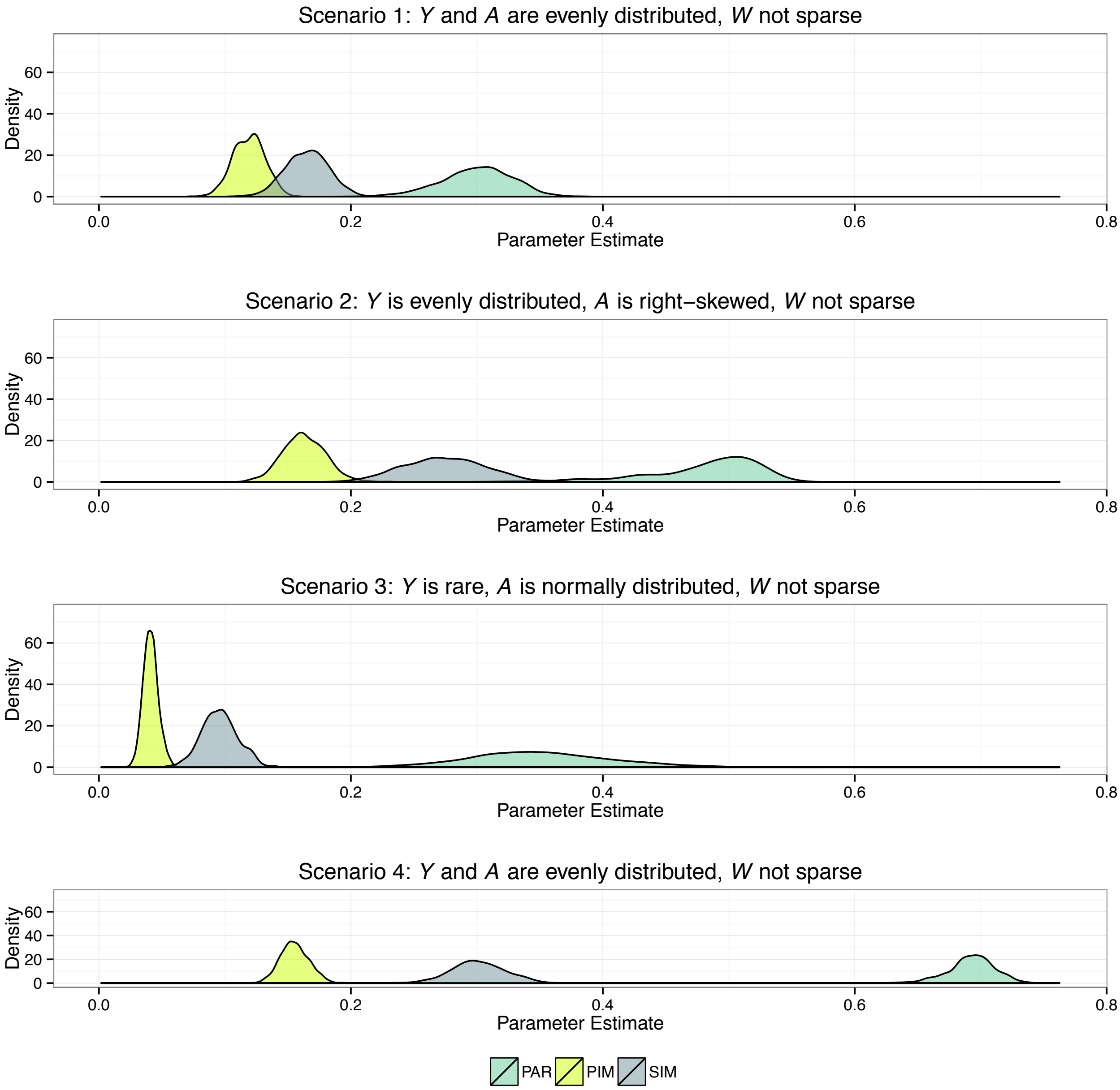


Figure C.3: Distribution of propensity scores for each scenario