

**Handwashing behavior and habit formation in the household:  
evidence from the pilot randomized evaluation of HOPE SOAP® in South Africa**



A minor dissertation submitted in partial fulfilment for the degree of  
Master of Commerce in Economic Development  
February 2018

**ABIGAIL SELLMAN**  
**SLLABI002**

School of Economics  
Faculty of Commerce  
University of Cape Town

Supervisors:  
Professor Justine Burns  
Dr Brendan Maughan-Brown

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

## ABSTRACT

Handwashing with soap at critical times is a simple and effective way to prevent the spread of communicable diseases, such as diarrhea and acute respiratory infection, which are major causes of morbidity and mortality in developing countries. However, rates of handwashing remain low throughout the world, and interventions which attempt to improve handwashing behaviors have largely been unsuccessful in practice. This may be because behavior change programs often fail to recognize the habitual drivers of handwashing behavior. In contrast, this paper examines the effectiveness of a novel soap technology, HOPE SOAP<sup>®</sup>, a child-size and colorful bar of soap with a toy embedded in its center, which aims to increase handwashing in children by specifically targeting its habitual nature. To rigorously evaluate HOPE SOAP<sup>®</sup>, this paper exploits data from a pilot randomized controlled trial whereby 229 households from a poor urban community in South Africa were randomly assigned to receive HOPE SOAP<sup>®</sup> for a period of 12-weeks. In an initial analysis of the effects of the intervention on children's health and behavior, Burns, Maughan-Brown, and Mouzinho (2017) found that that HOPE SOAP<sup>®</sup> had positive impacts on children's handwashing behaviors and health outcomes. Children who received HOPE SOAP<sup>®</sup> children were more likely to wash their hands, and had better overall health outcomes than control children (Burns, Maughan-Brown, and Mouzinho 2017).

Although HOPE SOAP<sup>®</sup> aims to induce behavior change in children, this paper explores the spillover effects that it has on other members of children's households. Specifically, this work uses regression analysis to investigate the impacts of HOPE SOAP<sup>®</sup> on the handwashing behaviors of children's primary caregivers, and on the health outcomes of all non-treated household members. This paper finds compelling evidence illustrating that a child's assignment to HOPE SOAP<sup>®</sup> has a positive impact on the handwashing behavior of their caregiver. Specifically, HOPE SOAP<sup>®</sup> increases the probability that a caregiver will wash their hands before eating a snack by 13 percentage points on average (p-value 0.17). A further investigation of the causal mechanisms for this improvement suggests that HOPE SOAP<sup>®</sup> affects caregiver behavior both by disrupting existing poor-hygiene habits, and by strengthening handwashing norms within households. Despite its positive effects on household handwashing behavior, this paper finds that a child's assignment to HOPE SOAP<sup>®</sup> has no discernable short-term impacts on the health of individual household members. Nevertheless, the positive influence of HOPE SOAP<sup>®</sup> on caregiver handwashing behavior is promising and, in conjunction with the finding that HOPE SOAP<sup>®</sup> improves children's behaviors, provides reason to believe the intervention may be successful in inducing habitual handwashing behaviors which can persist in the long run.

## **ACKNOWLEDGEMENTS**

This paper has not been previously submitted, in whole or parts, for the award of any other degree. All work and errors are my own. When I have relied on others' publications, I have made every attempt to properly cite and reference their works.

The pilot randomized controlled trial of HOPE SOAP<sup>®</sup>, on which this work is based, was supported by The Abdul Latif Jameel Poverty Action Lab (J-PAL) Incubation Fund. I gratefully acknowledge J-PAL for their funding of this project. I would also like to acknowledge all of the implementation and research staff who contributed to the evaluation's success at any stage. This paper would not have been possible without your initial efforts.

Finally, I would like to thank Professor Justine Burns and Dr Brendan Maughan-Brown for entrusting me with this project. Thank you for your invaluable support and wisdom throughout this process. It has been such a pleasure to work with you, and I look forward to our future collaborations. I would also like to thank my parents for teaching me to value education, and supporting me through all of my educational pursuits; as well as the rest of my family and friends, near and far, who have encouraged me on this journey.

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS .....	iv
LIST OF TABLES AND FIGURES .....	v
ABBREVIATIONS .....	vi
<b>1. INTRODUCTION .....</b>	<b>1</b>
<b>2. BACKGROUND.....</b>	<b>3</b>
2.1 HEALTH AND DEVELOPMENT .....	3
2.2 HANDWASHING FOR HEALTH .....	5
2.3 EVALUATING HANDWASHING PROMOTION PROGRAMS .....	7
2.4 UNDERSTANDING AND IMPROVING BEHAVIOR.....	9
2.5 SPILLOVERS FROM HEALTH INTERVENTIONS .....	12
2.6 CONTRIBUTIONS .....	14
<b>3. EXPERIMENTAL DESIGN .....</b>	<b>15</b>
3.1 HOPE SOAP <sup>®</sup> INTERVENTION .....	15
3.2 EVALUATION DESIGN.....	16
3.3 DATA COLLECTION .....	19
3.4 SAMPLE POPULATION.....	21
3.5 BASELINE HANDWASHING AND HEALTH .....	24
3.6 CHILD LEVEL RESULTS .....	27
<b>4. ANALYSIS AND RESULTS .....</b>	<b>28</b>
4.1 ESTIMATION STRATEGY .....	28
4.2 SELF-REPORTED VS. OBSERVED BEHAVIOR.....	31
4.3 RESULT 1: HOPE SOAP <sup>®</sup> IMPROVES CAREGIVER BEHAVIOR AT SNACK-TEST 2.....	32
4.4 RESULT 2: HOPE SOAP <sup>®</sup> IMPROVES CAREGIVER BEHAVIOR THROUGH TREATED CHILDREN .....	37
4.5 RESULT 3: HOPE SOAP <sup>®</sup> HAS DIFFERENTIAL EFFECTS, CONDITIONAL ON BASELINE CLEANLINESS .....	40
4.6 RESULT 4: HOPE SOAP <sup>®</sup> IMPROVES CAREGIVER RESPONSIVENESS TO THE FCW PROGRAM.....	42
4.7 RESULT 5: HOPE SOAP <sup>®</sup> 'S EFFECT IS NOT MEDIATED BY RESPONSIVENESS OR KNOWLEDGE.....	44
4.8 RESULT 6: HOPE SOAP <sup>®</sup> HAS NO EFFECT ON INDIVIDUAL HEALTH.....	46
4.9 SUMMARY OF RESULTS .....	48
<b>5. DISCUSSION.....</b>	<b>48</b>
5.1 LIMITATIONS AND CAVEATS .....	48
5.2 THE PROMISE OF HABITUAL BEHAVIOR CHANGE.....	51
5.3 LIMITED ABILITY TO OBSERVE HEALTH IMPACTS.....	54
<b>6. CONCLUSION .....</b>	<b>56</b>
REFERENCES .....	58
APPENDIX .....	63

## LIST OF TABLES AND FIGURES

### TABLES

1. SURVEY QUESTIONS AND RESPONSE CHOICES .....	20
2. HOUSEHOLD DESCRIPTIVE STATISTICS .....	22
3. CAREGIVER DESCRIPTIVE STATISTICS.....	23
4. INDIVIDUAL DESCRIPTIVE STATISTICS.....	24
5. CAREGIVER BASELINE HANDWASHING BEHAVIORS.....	25
6. INDIVIDUAL BASELINE HEALTH .....	26
7. MEANS OF CAREGIVER HANDWASHING BEHAVIOR .....	33
8. EFFECT OF HOPE SOAP <sup>®</sup> ON CAREGIVER BEHAVIOR (OLS RESULTS) .....	34
9. EFFECT OF HOPE SOAP <sup>®</sup> ON CAREGIVER BEHAVIOR (LOGIT/PROBIT RESULTS).....	36
10. CORRELATIONS BETWEEN CHILD AND CAREGIVER BEHAVIOR .....	38
11. CHILD BEHAVIOR AS A MEDIATOR FOR THE EFFECT OF HOPE SOAP <sup>®</sup> .....	39
12. DIFFERENTIAL EFFECTS OF HOPE SOAP <sup>®</sup> BY BASELINE CLEANLINESS .....	42
13. MEANS OF CAREGIVER RESPONSIVENESS VARIABLES.....	43
14. EFFECT OF HOPE SOAP <sup>®</sup> ON CAREGIVER RESPONSIVENESS TO THE FCW PROGRAM.....	44
15. INTEREST AND KNOWLEDGE AS ADDITIONAL COVARIATES.....	45
16. EFFECT OF HOPE SOAP <sup>®</sup> ON INDIVIDUAL HEALTH.....	47
A1. CHILD BEHAVIOR AS A MEDIATOR (LOGIT/PROBIT RESULTS) .....	63
A2. EFFECT OF HOPE SOAP <sup>®</sup> ON A SUBPOPULATION WITH UNCLEAN HOUSEHOLDS.....	65
A3. DIFFERENTIAL EFFECTS BY BASELINE CLEANLINESS (LOGIT/PROBIT RESULTS).....	66
A4. CAREGIVER RESPONSIVENESS SURVEY QUESTIONS .....	68
A5. INTEREST AS AN ADDITIONAL COVARIATE (LOGIT/PROBIT RESULTS) .....	69
A6. KNOWLEDGE AS AN ADDITIONAL COVARIATE (LOGIT/PROBIT RESULTS).....	71
A7. EFFECT OF HOPE SOPA ON INDIVIDUAL HEALTH (LOGIT/PROBIT RESULTS) .....	73
A8. EX-POST MDE AND SAMPLE SIZE CALCULATIONS.....	74
A9. EFFECT OF HOPE SOAP <sup>®</sup> ON HEALTH WITH A TOY-CHEAT INDICATOR .....	75

### FIGURES

F1. F-DIAGRAM AND DISRUPTIVE HWWS .....	6
F2. HOPE SOAP.....	15
F3. EXPERIMENTAL GROUPS.....	17
F4. EVALUATION TIMELINE .....	18
F5. INDIVIDUAL REPORTED HEALTH STATUS.....	27
F6. CAREGIVER REPORTED VS. OBSERVED BEHAVIOR.....	32
F7. CAREGIVER OBSERVED HANDWASHING BEHAVIOR .....	33

## ABBREVIATIONS

APE – Average partial effect

ARI – Acute respiratory infection

DALY – Disability adjusted life year

ECD – Early childhood development

FCW – Foundation for Community Work

FIF – Family-in-Focus

HWWS – Hand washing with soap

LPM – Linear probability model

MDE – Minimum detectable effect

OLS – Ordinary least squares

UCT – University of Cape Town

WASH – Water, sanitation, and hygiene

WHO – World Health Organization

WSP – Water and Sanitation Program (at the World Bank)

## 1. INTRODUCTION

Communicable diseases are a leading cause of morbidity and mortality in the developing world today. The most common of these diseases include diarrhea and acute respiratory infection (ARI), which are largely transmitted due to a lack of proper sanitation and hygiene. It is estimated that more than one-third of deaths of children under the age of five each year are caused by diarrhea and acute respiratory infection, with approximately 88% of diarrhea deaths attributable to poor sanitation and hygiene (World Bank 2005; Black, Morris, and Bryce 2003). While children bear the highest burden of disease due to improper sanitation and hygiene, adults' health can also be affected. Ill health may cause adults to miss work or be unable to complete simple day-to-day tasks, leading to a decline in productivity and loss of income (Rosen and Vincent 1999). In 2001 alone, diarrheal disease and acute respiratory infection accounted for 10% of all disability adjusted life years (DALYs) lost across the world (Lopez et al. 2006). Accordingly, population health and well-being are tightly linked with economic growth; and thus, promoting health through improving sanitation and hygiene is critical to a country's economic development and overall prosperity (Bloom and Canning 2008).

Handwashing with soap at critical times, such as before preparing food or after defecation, is a simple and effective way to improve hygiene and prevent the spread of diarrheal disease and acute respiratory infection (Luby et al. 2005; World Bank 2005; Briceño, Coville, and Martinez 2015). Handwashing with soap has been proven to decrease the incidence of diarrhea in children by up to 48% (Cairncross et al. 2010; Taylor et al. 2015). Nonetheless, rates of handwashing remain low throughout the world. In developing countries, it is estimated that only between 3-34% of individuals routinely wash their hands with soap at critical times (Galiani et al. 2016). Moreover, programs and interventions which attempt to increase handwashing practices through improving knowledge about the benefits of hand-hygiene have largely been unsuccessful in practice (Vindigni, Riley, and Jhung 2011). This may be because handwashing is a habitual behavior, driven largely by automatic processes and contextual cues (such as social norms), rather than knowledge or beliefs (Marteau, Hollands, and Fletcher 2012; Neal et al. 2015). Despite this, habit is poorly addressed in the literature on handwashing behavior change, and very few interventions which attempt to improve handwashing practices actually target the automatic drivers of habitual behavior (Curtis, Danquah, and Aunger 2009).

In contrast to the existing literature, this paper evaluates the effectiveness of HOPE SOAP<sup>®</sup>,<sup>1</sup> an innovative soap technology which aims to improve rates of handwashing by specifically leveraging insights into the habitual nature of the behavior. HOPE SOAP<sup>®</sup> is a child-sized, colorful, and translucent bar of soap with a toy inside of it, designed to make handwashing fun and engaging for children. The simple hypothesis is that HOPE SOAP<sup>®</sup> will induce children to wash their hands more regularly in order to obtain the toy inside the soap; and over time, repetitive handwashing behavior will lead to habit formation. In 2014, researchers Burns, Maughan-Brown, and Mouzinho from the University of Cape Town ran a pilot randomized controlled trial (229 households) to test for the impacts of HOPE SOAP<sup>®</sup> on children's handwashing behaviors and health outcomes.

---

<sup>1</sup> HOPE SOAP<sup>®</sup> is copyright of Young & Rubicam: <http://www.yr.com/>

The pilot was carried out in Delft, a poor-urban community outside of Cape Town in the Western Province of South Africa. An initial analysis of data collected during the pilot showed that HOPE SOAP<sup>®</sup> indeed had positive impacts on children's handwashing behaviors and health outcomes. When compared to a control group of children who simply received a bar of soap with a toy alongside of it, HOPE SOAP<sup>®</sup> children were more likely to wash their hands unprompted prior to eating a snack; more likely to wash their hands after using the toilet; and significantly more likely to use soap when washing their hands. HOPE SOAP<sup>®</sup> children also had better health outcomes in general than control children (Burns, Maughan-Brown, and Mouzinho 2017).

Results from the initial analysis certainly show that HOPE SOAP<sup>®</sup> had a positive impact on children's health and handwashing behaviors. However, little is known about the effects that the intervention had on the handwashing behaviors of children's caregivers, or on the health outcomes of other individuals in children's households. Understanding the indirect effects of HOPE SOAP<sup>®</sup> on non-targeted individuals is important for a few reasons. First, as handwashing is a habitual practice driven in part by social norms, it is plausible that improvements in children's behaviors may alter norms around handwashing within households; thereby causing caregivers to change their handwashing behaviors as well (Marteau, Hollands, and Fletcher 2012; Angelucci and Di Maro 2010). Furthermore, when the health of treated children improves, rates of disease transmission from these children to other individuals in close proximity to them will presumably decrease. This may lead to positive health benefits for all household members (Angelucci and Di Maro 2010). In interpreting the overarching impact of HOPE SOAP<sup>®</sup>, it is imperative to not only consider the intervention's effect on children, but also to account for potential spillovers to caregivers and household members. Quantifying the additional impacts (positive or negative) that a targeted intervention has on non-targeted individuals in the treatment context is critical to discerning the true effects of a program. For this reason, the importance of measuring spillovers is increasingly being recognized in the development and health economics literature, and numerous studies have solely focused on examining the indirect effects of health interventions (Chung et al. 2017). Nevertheless, to date, no evaluations exist which have explicitly examined spillovers in the context of a handwashing behavior change intervention (Chung et al. 2017).

Accordingly, the objective of this research paper is to investigate the effects that the HOPE SOAP<sup>®</sup> intervention, which targets children, has on the handwashing behaviors and health outcomes of other individuals residing in children's households. To explore the spillover effects of HOPE SOAP<sup>®</sup>, this paper exploits data collected from the aforementioned pilot randomized evaluation conducted by researchers at the University of Cape Town. By employing econometric estimation techniques, this paper generates evidence about the impacts of HOPE SOAP<sup>®</sup> on caregiver handwashing behavior and individual household-member health. Results from the empirical analysis, discussed at length below, demonstrate that a child's assignment to HOPE SOAP<sup>®</sup> treatment has a positive impact on the handwashing behavior of their caregivers. Specifically, a child's assignment to HOPE SOAP<sup>®</sup> improves the probability that a caregiver washes their hands when receiving a snack (during a second community health worker visit) by 13 percentage points on average (p-value 0.17), compared to caregivers whose children only receive normal soap. A further exploration of the causal

mechanisms for this improvement indicates that HOPE SOAP<sup>®</sup> affects caregiver behavior through two primary channels. First, HOPE SOAP<sup>®</sup> affects behavior by disrupting caregivers' existing poor-hygiene habits: the intervention has an especially strong and positive effect on caregivers who exhibit poor baseline cleanliness. Second, HOPE SOAP<sup>®</sup> may strengthen household social norms around handwashing. Evidence suggests that the correlation between the handwashing behaviors of household members is stronger for individuals in treatment households; and a further exploration of the causal mechanisms which lead to this difference implies that child behavior may mediate the treatment's effect on caregiver behavior. These findings are promising, and indicate that HOPE SOAP<sup>®</sup> may be successful in generating habitual handwashing behavior within households. However, despite the positive effects that the intervention has on handwashing behavior, this paper finds that HOPE SOAP<sup>®</sup> has little impacts on the health of individuals in treated households. Nevertheless, although HOPE SOAP<sup>®</sup> does not lead to any health improvements in the short-term, because it may successfully induce the habit of handwashing within households, in the long-run it could prove a useful way to decrease diarrheal disease and acute respiratory infection for individuals living in developing countries.

The remainder of this paper is structured as follows. First, Section 2 provides background information. The background section first explores the relationship between health, economic development, and sanitation and hygiene. It then provides a review of the existing literature on handwashing promotion interventions in developing country contexts. Next, Section 2 gives insights into the habitual determinants of handwashing behavior; and, lastly, it discusses the theory of spillovers in health interventions. Subsequently, Section 3 details the HOPE SOAP<sup>®</sup> intervention and the design of the pilot randomized evaluation. Section 3 also discusses the characteristics of sample households, and provides a summary of the results from the child-level analysis conducted by Burns, Maughan-Brown, and Mouzinho (2017). Section 4 first sets forth the estimation strategy used to analyze data on caregiver handwashing and household health. After describing the paper's approach to data analysis, Section 4 presents a detailed account of the results. Section 5 further discusses the results presented in Section 4; and includes insights into their limitations, policy implications, and contributions to the existing literature. Finally, Section 6 gives concluding remarks.

## **2. BACKGROUND**

### **2.1 Health and development**

Health and well-being are highly correlated with economic growth and prosperity (Bloom and Canning 2008; Clift 2004; Bleakley 2010). Due to an array of factors, individuals in rich countries tend to be healthier than individuals in poor countries. Reasons for this disparity include the fact that developed countries typically have better health care systems and sanitation infrastructures in place to support the health of their constituents; and additionally, individuals from high income countries are more likely to be able to afford investments into their own health (Bloom and Canning 2008). Just as better health is an outcome of living in a wealthy country, good

population health is also an input which leads to increased economic development. Similarly, poor population health causes economic stagnation and reinforces the problems of poverty – which, in turn, lead to worse health outcomes (Bleakley 2010). This cyclical relationship, between health and economic development, exists at both the macroeconomic level and the individual level. Historically, upward trends in macroeconomic growth have occurred in tandem with increases in life-expectancy (Bloom and Canning 2008). One estimate asserts that, on average, when a country gains an extra year of life-expectancy, it will also experience a 4 percent rise in per capita GDP (Bloom, Canning, and Sevilla 2004). Furthermore, as much as 50 percent of the economic growth differentials between developed and developing countries can be attributed to poor health and low life-expectancy (Collins 2015). On an individual level, health not only impacts an individual's well-being, but also affects their ability to reach their full economic potential (Bloom and Canning 2008).

Health primarily impacts economic development through its effects on human capital (Bloom, Canning, and Sevilla 2004). First, health influences labor productivity. Healthy individuals are more productive members of the workforce, and are able to participate in the workforce for longer (Bloom, Canning, and Sevilla 2004). At a household level, this directly translates into more earned-income (Bleakley 2010). Second, health affects educational attainment and investments into education (Bloom and Canning 2008). Healthy children are less likely to be absent from school, and more likely to stay in school for longer (Glewwe and Miguel 2007; Glewwe and Muralidharan 2015). Children who are healthier are also proven to earn higher incomes later in life (Glewwe and Miguel 2007). Moreover, when individuals know they will live for longer, they recognize greater returns to investments into education, and will thus pursue more education (Bloom, Canning, and Jamison 2004). Lastly, better health leads to heightened levels of savings and investments. In addition to having more income to save and invest, individuals who are healthier spend less money on health care, and can redistribute those finances to more productive activities, such as taking out a loan or starting a business. Additionally, when individuals expect to live longer, they also save more in preparation for their future (Bloom and Canning 2008; Clift 2004). In sum, the factors mentioned above contribute to, and reinforce, the vast inequalities in health and economic outcomes that exist between individuals living in high and low-income countries.

Not only do individuals from developing nations face worse health outcomes than individuals living in high-income countries, but they also encounter a different disease burden. In developed countries, the majority of deaths are caused by non-communicable diseases, such as diabetes or cancer, which tend to inflict older members of the population. For example, only 5% of all deaths in Europe are caused by infectious diseases (Curtis, Danquah, and Aunger 2009). In contrast, in developing countries, communicable diseases that disproportionately affect children cause the bulk of morbidity and mortality (Curtis, Danquah, and Aunger 2009; Dupas 2011). As of 2013, the under-five mortality rate in developing countries was a startling twelve-times higher than the rate in developed countries (United Nations Children's Fund 2014). HIV/AIDS and malaria are two major causes of child mortality in developing countries. However, an even larger number of children suffer from infectious diseases that are transmitted due to a lack of proper sanitation and hygiene. The most common of these diseases include diarrhea and acute respiratory infection (ARI) (World Health Organization 2009; Black, Morris, and Bryce 2003; Black et al. 2010). Diarrhea and ARI are often transmitted fecal-orally, when

bacteria and viruses are transferred from fecal matter to a new host, typically when the new host ingests contaminated water or food (World Bank 2005).<sup>2</sup>

The World Health Organization (WHO) estimates that one-third of child deaths around the world are as a result of diarrheal disease or ARI, with the majority of these deaths occurring in developing countries (World Health Organization 2009). In South Africa, where the child mortality rate is 52 per 1,000, 25.6% of all under-five deaths are as the result of diarrhea and ARI (World Health Organization 2017).<sup>3</sup> Furthermore, around 88% of deaths attributable to diarrhea across the world are as the consequence of unsafe water and/or a lack of access to adequate sanitation (Black, Morris, and Bryce 2003). Accordingly, diseases transmitted due to improper sanitation and hygiene also constitute for a huge loss in human capital and economic productivity. For example, in 1990, at least 8% of all disability-adjusted life years (DALYs) lost in the developing world were due to diarrhea alone; and in 2001, diarrheal disease and ARI accounted for 10% of all DALYs lost across the world (Lopez et al. 2006; Zwane and Kremer 2007).<sup>4</sup> Hence, diarrheal disease and ARI result in massive lost opportunities for economic growth, and as such, investments into improving water, sanitation, and hygiene should be a priority for developing countries.

## **2.2 Handwashing for health**

The World Bank estimates that as much as 90 percent of diarrhea, and 15% of under-5 diarrhea deaths, could be prevented by improving water, sanitation, and hygiene (WASH) in the developing world (World Bank 2005). Making improvements to WASH infrastructures, for example by building latrines or piped-water systems, is critical for decreasing rates of diarrhea and ARI. However, investing in infrastructural projects is often infeasible for developing countries today. Large-scale infrastructure improvements typically require vast amounts of capital (i.e. financial resources, labor, etc.) at inception, need management and upkeep over time, and can still leave proportions of the population without adequate access (Zwane and Kremer 2007). Furthermore, even when proper sanitation infrastructures do exist, poor individual health outcomes often persist because human behavior is also a key determinant of diarrheal disease and ARI (Marteau, Hollands, and Fletcher 2012). As such, efforts to improve WASH related health outcomes in developing countries should not only focus on constructing better infrastructures, but also, on improving individual hygiene behaviors.

Handwashing with soap (HWWS) is a simple, effective, and inexpensive way to prevent the spread of disease due to fecal matter contamination. This is because handwashing with soap at critical times, such as before preparing food or after defecation, disrupts the transmission path of a virus or bacteria and restricts its ability to infect a new host (Luby et al. 2005; World Bank 2005; Briceño, Coville, and Martinez 2015). The “F-

---

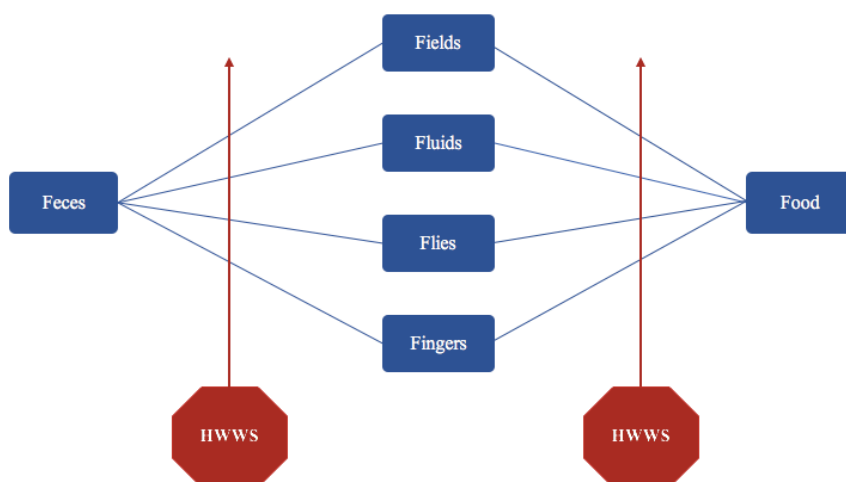
<sup>2</sup> Feces are one of the largest carriers of disease: one gram of feces can contain up to 1 million bacteria and 10 million viruses (World Bank 2005).

<sup>3</sup> This under-5 mortality rate is significantly lower than the average rate for the rest of sub-Saharan Africa (92 deaths per 1,000), but much higher than the developed country average (6 per 1,000) (United Nations Childrens Fund 2014).

<sup>4</sup> DALYs are an indicator which reflect the total amount of healthy life lost from a disease, either by mortality or disability (World Bank 2005). DALYs are designed to place value on an individual's life, and take into account the amount of possible productivity lost due to a disease (Hyder, Puvanachandra, and Morrow 2012). In this way, DALYs provide a comparable way to assess the total burden of diseases on a society (World Bank 2005).

Diagram” presented in Figure 1 below shows how handwashing with soap (HWWS) at critical times works to prevent the spread of illness from fecal matter to food (and/or water) by disrupting the path of disease transmission.<sup>5</sup>

**Figure 1: F-diagram and disruptive HWWS**



Moreover, because it directly disrupts the path of disease transmission at an individual level, handwashing has the potential to be effective in reducing infectious disease even in the absence of basic sanitation: for instance, in locations which have poor infrastructures and high rates of environmental fecal contamination (Orsola-Vidal and Yusuf 2011; World Bank 2005). In addition to being successful in preventing disease, promoting handwashing behavior is also relatively inexpensive and therefore highly cost-effective. Multiple studies cite handwashing with soap as one of the most cost-effective ways to prevent the spread of infectious disease (Cairncross and Valdmanis 2006; World Bank 2005; Borghi et al. 2002).

Despite its widely-recognized benefits, rates of handwashing remain low throughout the world: only between 3-34% of individuals in developing countries routinely wash their hands at critical times (World Bank 2005). Furthermore, there are vast differences in the handwashing practices between individuals living in developed and developing countries. For instance, one retrospective study using data from 145 countries estimates that the mean rate of handwashing after fecal contamination is 43% in high-income countries, but only 14% in low and middle-income countries (Prüss-Ustün et al. 2014).<sup>6</sup> Thus, because of the current low rates of handwashing, improving the handwashing behaviors of individuals in developing countries has the potential to drastically reduce the spread of diarrhea and ARI.

<sup>5</sup> Adapted from Briceño, Coville, and Martinez (2015) and World Bank (2005).

<sup>6</sup> Prüss-Ustün et al. (2014) estimate that the global mean rate of handwashing (including both high and low-middle income countries) is 19%.

### **2.3 Evaluating handwashing promotion programs**

Although, in theory, promoting handwashing holds great promise for reducing the burden of disease in developing countries, limited research exists into how handwashing with soap actually affects individual health outcomes in practice. A few recent studies on the effectiveness of handwashing indicate that handwashing can prevent anywhere from 23 to 48 percent of diarrhea cases, and 23 to 30% of respiratory infections in children (Rabie and Curtis 2006; Curtis and Cairncross 2003; Curtis, Danquah, and Aunger 2009; Cairncross et al. 2010; Ejemot et al. 2008; Luby et al. 2005; Prüss-Ustün et al. 2014). However, most of these assessments were not conducted in contexts particularly relevant to public health; but rather, in institutional settings, under tightly controlled trial conditions, or with consistent involvement by investigators (Chase and Do 2012; Aunger et al. 2010).<sup>7</sup> Because of this, in recent years, field research into the effects of handwashing promotion programs in low-income contexts has become a priority for many development organizations, and funding for such work has become more widely available (Schmidt 2014). For instance, in 2006 the Water and Sanitation Program (WSP) at the World Bank created the Global Scaling Up Handwashing program. The program aimed to increase handwashing behavior in developing countries, and consisted of multifaceted large-scale handwashing-promotion programs implemented in multiple countries such as Peru, Senegal, Tanzania, and Vietnam. In order to produce further knowledge about the effects of handwashing with soap in developing country contexts, the WSP also executed rigorous impact evaluations in tandem with program implementation (Chase and Do 2012).

The handwashing programs implemented by the World Bank generally consisted of intervention packages which contained multiple handwashing-promotion elements. Some components of the interventions included: mass-media campaigns to increase knowledge about the benefits of handwashing at key times; commercial marketing schemes to encourage the purchase of soap products; and handwashing trainings for community activists which aimed to stimulate community-wide behavior change (Chase and Do 2010; Orsola-Vidal and Yusuf 2011; Galiani, Gertler, and Orsola-Vidal 2012; Briceño, Coville, and Martinez 2015). Despite the considerable variety of tactics utilized by the WSP, evidence generated by the impact evaluations on the impacts of promoting handwashing remains inconclusive. Results show that the interventions had mixed effects overall on handwashing behaviors and health outcomes. For example, in Tanzania, Peru, and Vietnam, the WSP interventions were successful at improving caregivers' knowledge of proper handwashing techniques and critical times (Chase and Do 2012; Galiani et al. 2016; Briceño, Coville, and Martinez 2015). However, these improvements in knowledge often did not translate into better handwashing behaviors or health outcomes. In Vietnam and Tanzania, there were no significant differences in the observed handwashing behaviors of individuals who received the interventions as compared to individuals in control groups who did not receive the interventions. Accordingly, individuals in neither study enjoyed health benefits as a result of the handwashing promotion programs (Chase and Do 2012; Briceño, Coville, and Martinez 2015). In Peru however, the intervention did lead to a large positive impact on handwashing behavior. Individuals who received the intervention package were 61% more likely to wash their hands before eating, and 69% more likely to wash

---

<sup>7</sup> For other notable examples, see Rabie and Curtis (2006) and Ejemot et al. (2008).

their hands before preparing food than individuals in a control group (Galiani et al. 2016). Nevertheless, in Peru, the intervention did not have any positive impacts on health outcomes: there were no significant differences in rates of diarrhea or ARI between children under-5 in the treatment and control groups (Galiani et al. 2016).

Similarly, a review of additional literature on the effects of handwashing encouragement programs in developing countries also reveals mixed results. For example, a randomized evaluation by Biran et al. (2009) found negative impacts of a handwashing program in India. Like the packages implemented by the WSP, the intervention assessed by Biran et al. (2009) consisted of an assortment of hygiene promotion activities, including: school visits, structured meetings with community leaders, and public addresses aimed at informing mothers about the presence of germs on hands. Surprisingly though, Biran et al. (2009) discovered that handwashing behavior actually declined as a result of the program, despite the fact that individuals who received the intervention had higher levels of self-reported germ awareness. Specifically, post-intervention, adults in communities that received the encouragement package washed their hands at critical times approximately 3 percentage points less than they did prior to receiving the intervention (Biran et al. 2009). A second study by Cairncross et al. (2005) also evaluated a hygiene promotion program in India. Unlike Biran et al. (2009) however, Cairncross et al. (2005) found positive impacts of a handwashing intervention on handwashing behavior. More than half of the adults who received the handwashing encouragement intervention had greater knowledge about proper handwashing techniques, and reported handwashing at critical times significantly more than individuals in a control group.

A third study, conducted in Pakistan, found that a handwashing intervention which provided information on handwashing, coupled with weekly field worker visits and soap distribution, had positive impacts on child health (Luby et al. 2005). In a randomized evaluation of the aforementioned intervention, Luby et al. (2005) found that children in communities who received the intervention package had significantly better health outcomes than children in communities who did not receive the intervention. Specifically, children from treated communities experienced 50% fewer instances of cough, congestion, and pneumonia (symptoms and illnesses related to ARI) than children from communities who did not receive the handwashing promotion package (Luby et al. 2005). However, this positive health finding is somewhat of an outlier, for most studies (like Biran et al. (2009) mentioned above) have failed to find a link between handwashing promotion programs and improvements in handwashing behavior or health in the long-term. In a systematic review of 30 articles that evaluate handwashing behavior change, Vindigni, Riley, and Jung (2011) concluded that nearly every one of the studies reported that handwashing promotion programs led to some positive behavior changes in the short-term. However, none of the studies definitively proved lasting behavior change in the long run, and the majority of the positive behavioral impacts found did not translate to improvements in health (Vindigni, Riley, and Jung 2011).

Clearly, the literature is inconclusive, and there is much uncertainty about the ability of handwashing encouragement programs to positively impact the handwashing behaviors and subsequent health outcomes of individuals living in developing countries (Vindigni, Riley, and Jung 2011). Furthermore, even when the

hygiene promotion programs delineated above were found to be successful at creating positive behavior change, the evaluated-interventions themselves are not particularly policy relevant. In practice, comprehensive package interventions, such as those evaluated above, are likely too costly and require too many scarce resources to be implemented by developing countries in the long-run or at scale (Zwane and Kremer 2007). Hence, more research is needed to distinguish which specific components of handwashing encouragement programs are most cost-effective and successful in generating handwashing behavior change.

Moreover, despite the mixed findings presented above, there is one trend which was present throughout much of the literature. Handwashing promotion programs almost always led to improvements in individual knowledge about proper handwashing techniques (i.e. Chase and Do 2012; Galiani et al. 2016; Briceño, Coville, and Martinez 2015; Biran et al. 2009; Cairncross et al. 2005), but better knowledge only seldom translated into actual handwashing behavior change (i.e. Chase and Do 2012; Galiani et al. 2016; Briceño, Coville, and Martinez 2015; Biran et al. 2009). This discrepancy, between knowledge and action, is troubling and warrants a deeper exploration into the behavioral determinants of handwashing with soap.

#### **2.4 Understanding and improving handwashing behavior**

As exemplified in the literature above, there is often a gap which exists between an individual's knowledge about the benefits of handwashing and their actual handwashing behaviors. Insights from behavioral science establish that this gap may largely be due to the habitual nature of handwashing. Habits are behaviors which are learned incrementally over time; become automatic when they are repeated many times over; and, once formed, are triggered by contextual cues rather than motivated by knowledge and beliefs (Neal et al. 2015; Aunger et al. 2010). Some estimates claim that as much as 50% of humans' daily activities are habitual (Neal et al. 2015; Curtis, Danquah, and Aunger 2009). Because handwashing occurs as part of a daily routine, it is repeated frequently enough to become habitual (Aunger et al. 2010; Curtis, Danquah, and Aunger 2009). Similarly, even when individuals do not engage in handwashing behaviors, they are still regularly presented with opportunities in which handwashing should take place – and therefore can form a habit of not handwashing. In addition to its repetitive nature, handwashing is habitual because it is largely driven by contextual cues such as the presence of soap, key occasions (i.e. defecation, food preparation), and social norms (Aunger et al. 2010; Curtis, Danquah, and Aunger 2009; Neal et al. 2015).

In a factor analysis on the determinants of handwashing behavior in Kenya, Aunger et al. (2010) found that habit was the most significant predictor of handwashing with soap. Further, Aunger et al. (2010) also reported that individuals themselves recognized the influence that habits had on determining their handwashing behaviors. Individuals who reported that they had an established habit of handwashing were significantly more likely to actually engage in handwashing behavior.<sup>8</sup> Additionally, a recent study in India found that individuals who knew their handwashing practices would be monitored and incentivized in the future

---

<sup>8</sup> Handwashing behavior in Aunger et al. (2010) was measured through direct observation: interviewers observed the handwashing behaviors of household members at critical times for a period of three hours.

increased their handwashing behaviors in the present, in an attempt to solidify a habit of handwashing before monitoring began (Hussam et al. 2016). This again suggests that individuals recognize the habitual nature of their own handwashing behaviors.

In contrast to reflective or planned behaviors, which are goal-oriented and based on beliefs and intentions, habitual behaviors are carried out automatically and require little cognitive thought (Neal et al. 2015; Aunger et al. 2010). However, habitual behaviors do typically begin as actions that are performed intentionally (Ouellette and Wood 1998). Initially, goals and intentions help to establish behaviors; and then, once behaviors are routinely practiced, they continue as habits largely determined by automatic processes (Neal et al. 2015). Thus, attitudes and intentions help to form new behaviors (and influence those performed infrequently), but have little power over behaviors once they are habitual. Instead, past-behaviors and contextual cues are the primary drivers of current habitual behaviors (Ouellette and Wood 1998; Neal et al. 2015; Curtis, Danquah, and Aunger 2009).

Although handwashing is widely recognized as a habit, habit is poorly addressed in the literature on handwashing behavior change, and very few interventions that attempt to change handwashing behavior target habitual (or automatic) processes (Curtis, Danquah, and Aunger 2009). As exemplified in the literature review in Section 2.3, most interventions aimed at improving handwashing behavior focus on delivering information and education on the benefits of handwashing. Information interventions encourage individuals to consider the motivations behind their actions, and make cognizant changes to their behaviors: implicitly assuming that behaviors are reflective and planned (Marteau, Hollands, and Fletcher 2012). However, as shown, this is not the case in the context of habitual behaviors; and, accordingly, interventions which concentrate on changing beliefs or intentions are likely ineffective at inducing behavior change in handwashing with soap (Marteau, Hollands, and Fletcher 2012; Curtis, Danquah, and Aunger 2009).<sup>9</sup>

Unsuccessful interventions which target habitual behaviors like handwashing fail for two primary reasons. First, interventions may fail to disrupt existing habits, such as an engrained habit of not-handwashing (Neal et al. 2015). Failure to disrupt existing habits may be one reason that, while many handwashing encouragement programs succeed at improving knowledge of handwashing behaviors, this knowledge does not translate into increased handwashing with soap. The second reason programs which target habits fail is because even when interventions are successful in changing behaviors at first, they may fail to get new behaviors to stick. Old habits tend not to be forgotten, and individuals often relapse back into old habits over time, even if they engage in a new behavior at first (Neal et al. 2015). This may help to explain why, in their systematic review of research on handwashing interventions, Vindigni, Riley, and Jhung (2011) found that although many handwashing promotion programs led to behavior change in the short-term, none could prove lasting change in the long-run.

---

<sup>9</sup> This problem, of unsuccessful education/information interventions, not only exists for programs which target handwashing, but also in an array of other preventive health contexts in both developed and developing countries (see Marteau, Hollands, and Fletcher (2012) and Dupas (2011) for notable examples).

Despite these challenges, crafting programs to change habits is not impossible; but simply requires employing insights into the psychological drivers of habitual behaviors, and creating interventions which primarily target automatic, rather than intention driven, processes (Neal et al. 2015; Marteau, Hollands, and Fletcher 2012). Two notable articles written by behavioral scientists provide some concrete suggestions on how to best craft interventions in the circumstance of habitual behaviors. Neal et al. (2015) and Marteau, Hollands, and Fletcher (2012) note that the following techniques, among others, may help programs succeed in changing habitual behaviors. First, programs should ensure environmental contexts are conducive to supporting habitual behavior change (Neal et al. 2015; Marteau, Hollands, and Fletcher 2012). In the instance of handwashing, this could mean ensuring that households have a designated handwashing station and access to soap (Neal et al. 2015). Second, interventions should leverage contexts or use "teachable-moments" when attempting to disrupt existing habitual behaviors (Neal et al. 2015; Marteau, Hollands, and Fletcher 2012). A teachable moment refers to an opportunity in time, or a set of circumstances, which implicitly encourages individuals to change their behaviors (Lawson and Flocke 2009). For example, the birth of a child constitutes as a teachable moment; and using the birth of a child as a time to teach mothers about the health benefits of handwashing may help programs succeed in promoting handwashing behavior change (Neal et al. 2015). Third, programs which hope to change habitual behaviors should try to eliminate choice frictions or constrain choices to help automatically direct behaviors towards specific actions (Marteau, Hollands, and Fletcher 2012). Fourth, programs should attempt to provide motivations for actions that focus on using automatic, rather than cognitive, processes to change behaviors (Neal et al. 2015).

One way interventions can provide automatic motivators for behaviors is by giving rewards, or other positive reinforcements, for completing specific actions (Neal et al. 2015). Rewards and incentives make behaviors goal-oriented, and motivate behavior change without requiring individuals to think about the beliefs and intentions behind their actions. In addition to making behaviors goal-oriented, rewards give actions a positive connotation. Humans are predisposed to seek out positive stimulation (and avoid negative stimulation); and therefore, rewards can inadvertently motivate behavior change by tapping into this psychological trait (Marteau, Hollands, and Fletcher 2012). Even without rewards or incentives, positive associations alone can help motivate behavior change. For instance, something as simple as using fun words to describe healthy foods has been proven to actually increase the likelihood that children will eat them (Marteau, Hollands, and Fletcher 2012). Crafting interventions which give actions positive associations or make them goal-oriented help programs overcome one of the bottlenecks to habitual behavior change – failure to disrupt existing habits. Furthermore, to overcome the second major obstacle in creating habitual behavior change – the challenge of generating behavior change that sticks – behavioral scientists suggest that programs should ensure that they encourage repetitive practice as part of introducing a new action (Neal et al. 2015). Through repetition, new behaviors can become habitual themselves; and repetition helps to ensure individuals do not relapse into old behavioral patterns over time. Section 3.1 below details how HOPE SOAP<sup>®</sup>, whose effects on handwashing and health are evaluated as the focus this paper, leverages the above insights into the behavioral determinants

of handwashing to create a soap product which aims to generate lasting habitual behavior change in handwashing with soap.

## **2.5 Importance of evaluating spillovers in health interventions**

As expounded upon above, although it is a habitual behavior, habit is poorly addressed in the literature on handwashing, and few interventions target the habitual nature of handwashing when attempting to improve behaviors. Filling this gap in the literature could have sizable implications for handwashing behavior and health in developing countries; and thus, further research into interventions which explicitly target the habitual nature of handwashing is warranted. A second, and equally as important, area of study that deserves more attention with respect to handwashing in low and middle-income countries is health-related spillovers. Spillovers can be defined as the indirect effects of a program or intervention on a non-treated population, caused due to the non-targeted group's interactions with the treated group (Angelucci and Di Maro 2010). Poverty alleviation, behavior change, and health improvement interventions typically target a specific population. However, because individuals consistently engage in economic and social-behavioral interactions with those around them, spillover effects almost always occur (Angelucci and Di Maro 2010). Spillovers are likely to occur even when interventions have a limited scope and implementation protocols are tightly controlled and monitored.

In relation to economic development research focused on health, spillovers can typically be grouped into three distinct categories. These include: (1) interactions or substitutions, (2) context equilibrium effects or social proximity effects, and (3) health externalities or disease transmission effects (Angelucci and Di Maro 2010; Chung et al. 2015).<sup>10</sup> Interactions and substitutions describe effects on the non-target population that occur through their social and economic exchanges with the treated group (Angelucci and Di Maro 2010; Chung et al. 2015). For example, in the context of handwashing promotion programs, interaction effects may occur if a treated group shares knowledge they learn about proper handwashing practices with a non-treated population; and individuals in the non-treated population update their behaviors accordingly. Furthermore, a substitution effect could occur if, for instance, a household-head receives a free bar of soap and is able to reallocate the money they would have used to purchase the soap towards (for example) additional food for their children; hence improving children's nutritional status.

Context equilibrium effects or social proximity effects refer to the behavioral implications that an intervention has for non-treated members of the population. Interventions produce behavioral spillovers when treated individuals change their behaviors as the result of an intervention, and subsequently incite behavior change in non-treated individuals (Chung et al. 2015). Behavioral spillovers can occur between individuals who live in close proximity to one another (i.e. siblings, neighbors, etc.), or because an intervention changes

---

<sup>10</sup> Angelucci and Di Maro (2010) also list (4) general equilibrium effects as an important spillover category. They define general equilibrium effects as: "...effects that an intervention, which targets only part of the 'local economy', can have on the entire population. Active labor market policies or any intervention that can affect equilibrium prices are examples of these types of spillover effects" (Angelucci and Di Maro 2010). General equilibrium effects are less relevant to this research, as this paper focuses on measuring changes in individual outcomes rather than market outcomes.

the social-behavioral norms surrounding a particular behavior in a broader context (Angelucci and Di Maro 2010). Both of these transmission pathways (proximity or social norms) could occur as a result of a targeted handwashing intervention. For instance, non-treated siblings of children who are provided with a handwashing intervention in school may update their handwashing behaviors to mimic the new handwashing behaviors of their treated siblings. Furthermore, a handwashing promotion package could succeed in generating handwashing behavior in specific households within a community, and eventually, over time, the entire community's social norms could shift towards better handwashing. Moreover, as social norms are a key determinant of habitual behaviors (because they act as a contextual cue), behavioral spillovers that change social norms may be especially relevant to handwashing.

The last type of spillover effects which are applicable to research on handwashing are health externalities (or disease transmission effects). Health externalities occur when an intervention decreases the instance of illness in a treated group, thereby lowering disease transmission rates across an entire population. One commonly cited example of health externalities is a paper on school-based deworming by Miguel and Kremer (2004). In their evaluation of a deworming program in Kenya, Miguel and Kremer (2004) found that providing deworming drugs to specific children had an indirect effect on lowering infection rates for non-treated children as well, due to lower transmission of worms between all children within schools. Furthermore, as parasites negatively affect children's school attendance and performance, non-treated children also enjoyed better educational outcomes as a result of the intervention (Miguel and Kremer 2004). A similar situation could hypothetically occur in the context of handwashing: improving the health of treated individuals through handwashing with soap could lead to less disease transmission, and therefore less instance of diarrhea and ARI for non-treated individuals within their networks.

Accounting for intervention spillovers when conducting impact evaluations is important for a few primary reasons. First, accounting for spillovers is necessary to accurately measure the true effect of a program. If an intervention has influence on a non-treated population that is not measured, approximations of the intervention's effects will misrepresent the true impacts of a program (Angelucci and Di Maro 2010; Chung et al. 2017). All of the examples provided above allude to positive spillover effects, or those that work in the same direction as the intended treatments (i.e. improving health for treated and non-treated individuals). Failure to account for positive spillovers leads to underestimating the impacts of a treatment. However, in practice, negative spillovers are just as likely to occur. Negative spillovers can dilute or reverse an intervention's effects, and cause the true effect of a program to be overestimated (i.e. poor health in non-treated individuals may lessen the positive impact of a program on the treated population). If either positive or negative spillovers are not accounted for when analyzing the effects of a program, not only will treatment estimates give biased results, but they also may lead researchers to make the wrong policy conclusions regarding a specific program (Angelucci and Di Maro 2010; Chung et al. 2015). Moreover, measuring spillover effects is an interesting and valuable research topic in itself (Chung et al. 2017). Just as understanding the determinants of human behaviors can lead to more carefully crafted interventions (as described in Section 2.4), understanding spillover mechanisms can also help lead to the creation of better and more cost-effective programs. Thus, by building

evaluations which explicitly measure spillovers, researchers and policy makers will gain insights into the mechanisms which cause interventions to produce outcomes: leading to the development of better policies and programs.

For these reasons, including analyses of spillover effects is becoming more routine in the development economics literature today (Angelucci and Di Maro 2010; Chung et al. 2017). A recent systematic review published by Chung et al. (2017) discovered that 54 articles which specifically focused on measuring the spillover effects of health interventions were published between 2010 to 2013. However, the systematic review did not find any research articles which specifically measured spillovers from handwashing interventions; and only found three evaluations which studied spillovers in relation to water and sanitation programs more generally (Chung et al. 2017). Thus, because of the lack of research into spillovers of WASH interventions, combined with the likely possibility that indirect effects do occur in the context of sanitation and hygiene, much further research is needed. Accordingly, the evaluation presented in this paper attempts to contribute to the spillover literature by analyzing the indirect impact of HOPE SOAP<sup>®</sup>, a soap product targeted at children, on the behaviors of children's primary caregivers and the health outcomes of other members in children's households. Section 3, forthcoming, will provide information on the HOPE SOAP<sup>®</sup> intervention itself; insights into the causal chains by which spillovers from HOPE SOAP<sup>®</sup> may occur; and a description of the data this paper will use to test for the presence of spillover effects.

## **2.6 Contributions**

In sum, Sections 2.3, 2.4, and 2.5 identify significant gaps in the existing literature on handwashing interventions in developing countries. Firstly, although several published papers illustrate a positive link between handwashing and health, few of these studies occurred in low-income countries, or in settings particularly relevant to public health in general. The rigorous evaluations of handwashing programs which did occur in relevant contexts provide mixed results on average: the evidence is varied on whether or not handwashing encouragement programs can be effective at improving handwashing behaviors and reducing diarrhea and ARI in low-income countries. While many handwashing interventions did lead to positive improvements in knowledge about handwashing, increased knowledge seldom translated into better behaviors. In addition, even when handwashing encouragement programs were successful at changing behaviors in the short-term, there is no evidence that these changes were sustained in the long-run. This may be due to the fact that handwashing is a habitual behavior. Despite the fact that handwashing is widely recognized as a habitual behavior, habit is poorly addressed in the literature on handwashing behavior change; and few programs have leveraged this knowledge to create successful handwashing interventions. Furthermore, the types of interventions which have been evaluated in the literature are typically large, comprehensive handwashing promotion programs. As such, even when these programs were proven successful through rigorous evaluation, the evidence may be irrelevant for policy makers because package interventions are unrealistic to implement in practice. Lastly, although it is quite plausible to assume that behavioral spillovers and health externalities occur in the context of handwashing, no studies which explicitly measure the indirect effects of handwashing interventions exist to date.

Accordingly, this paper contributes to the literature in a few major ways. Firstly, by interrogating the behavior spillovers and health externalities that may exist due to HOPE SOAP<sup>®</sup>, this paper generates much needed evidence on spillovers which occur from WASH and handwashing behavior change interventions. Second, this paper evaluates HOPE SOAP<sup>®</sup>, a novel soap technology which aims to improve handwashing by leveraging insights into the habitual nature of handwashing behavior: adding breadth to the existing literature on habit and handwashing. Finally, HOPE SOAP<sup>®</sup> is a single, simple technology which could be easily implemented in practice; and as such, results from this evaluation may prove useful for policy makers who hope to design a handwashing intervention which would successfully induce habitual behavior change, and could realistically be implemented in developing country contexts.

### 3. EXPERIMENTAL DESIGN

#### 3.1 HOPE SOAP<sup>®</sup> intervention

HOPE SOAP<sup>®</sup> is a child-sized, colorful, and translucent bar of glycerin soap with a toy embedded in its center. HOPE SOAP<sup>®</sup> was jointly developed by the organizations Young & Rubicam (Y&R), a marketing and communications firm, and Safety Lab, a non-profit behavioral innovation lab. HOPE SOAP<sup>®</sup> was designed with the primary objective of instilling the habit of handwashing children, and the ultimate goal of improving children's health outcomes (Burns, Maughan-Brown, and Mouzinho 2017). HOPE SOAP<sup>®</sup>'s design embodies insights from behavioral science to encourage handwashing habit formation in children in a few ways. First, the toy placed in the center of HOPE SOAP<sup>®</sup> makes handwashing fun and goal-oriented (Burns, Maughan-Brown, and Mouzinho 2017). In this way, HOPE SOAP<sup>®</sup> encourages children to engage in handwashing behavior. Furthermore, by making handwashing behavior goal-oriented, HOPE SOAP<sup>®</sup> may also be able to disrupt children's existing poor-hygiene habits. Second, because children must wash their hands many times in order to obtain the toy at the soap's center, HOPE SOAP<sup>®</sup> ensures that the newly incited behavior is repeated consistently. By encouraging repetition in handwashing behavior, HOPE SOAP<sup>®</sup> may be able to instill a new habit of handwashing in children: generating behavior change that persists (or sticks) over time.

Figure 2: HOPE SOAP<sup>®</sup>



### 3.2 Evaluation design

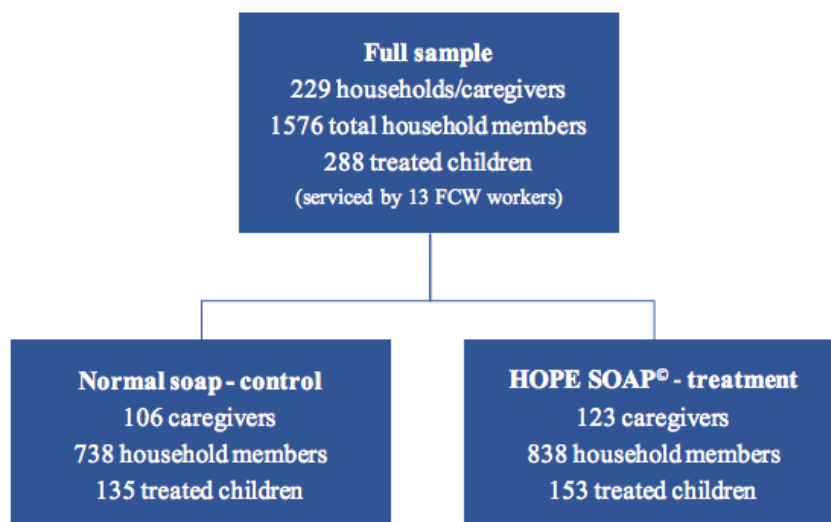
While, in theory, HOPE SOAP<sup>®</sup> could help children to develop a habit of handwashing, this may or may not be the case in practice. To evaluate if HOPE SOAP<sup>®</sup> can actually be successful in inciting habitual handwashing behavior, in 2014 Burns, Maughan-Brown, and Mouzinho (2017) carried out a pilot impact evaluation of the intervention. The evaluation was conducted in the poor urban community of Delft, outside of Cape Town in the Western Province of South Africa. The impact evaluation was completed in partnership with the Foundation for Community Work (FCW), an early childhood development (ECD) non-profit organization. Foundation for Community Work runs the Family-in-Focus (FIF) program in Delft and multiple other communities in the Western Province. Households who are enrolled in the Family-in-Focus program receive bi-weekly visits from trained community health workers. During these visits, health workers engage with children's caregivers on a variety of ECD issues (including health and hygiene), and help to facilitate child-caregiver interactions (Burns, Maughan-Brown, and Mouzinho 2017). In 2013 alone, Foundation for Community Work's Family-in-Focus program served over 10,000 families and 13,000 children. In Delft, 13 community health workers serviced households enrolled in the Family-in-Focus program (Burns, Maughan-Brown, and Mouzinho 2017). In order to evaluate the impacts of HOPE SOAP<sup>®</sup> on handwashing behavior and subsequent health outcomes, HOPE SOAP<sup>®</sup> was distributed to households in the Foundation for Community Work's program as an addition to the typical Family-in-Focus health and hygiene curriculum.

The pilot evaluation of HOPE SOAP<sup>®</sup> was designed as a randomized controlled trial (RCT) (Burns, Maughan-Brown, and Mouzinho 2017). An RCT, or randomized evaluation, is a rigorous research method designed to allow for the accurate assessment of the effects of a program or intervention on a population (Gertler et al. 2011; Glennerster and Takavarasha 2013). In the fields of development economics and public health today, RCTs are often considered the gold-standard of research methodologies (Gertler et al. 2011). RCTs work by randomly assigning members in a population of interest to different experimental groups, who receive distinct interventions. Typically, one experimental group does not receive any intervention, and therefore acts as comparison, or control group, for researchers to use as a benchmark when assessing the impact of a program on a treated group. Baseline measures of demographic variables and outcomes-of-interest are taken for the whole study population; and upon confirmation that members of the different experimental groups have equal characteristics at the onset of a study, the comparison group can act as a valid counterfactual for the treatment (Gertler et al. 2011; Glennerster and Takavarasha 2013). In other words, when individuals in the treated and control groups are identical at baseline, random assignment allows researchers to assume that, in the absence of an experiment, both groups would have identical outcomes over time. This allows researchers to interpret the effects of a program on the treatment group as causal (Orsola-Vidal and Yusuf 2011; Duflo, Glennerster, and Kremer 2006).

To be eligible for participation in the randomized evaluation of HOPE SOAP<sup>®</sup>, households in the Family-in-Focus program must have met the following three criteria. (1) Households must have had at least one child aged 3 to 9 years old to receive the soap. (2) The primary caregivers must have expressed interest in continued Family-in-Focus program participation. (3) Children must not have been enrolled in any other ECD

program (Burns, Maughan-Brown, and Mouzinho 2017).<sup>11</sup> In total, 229 households (288 children) in Delft were eligible to participate in the study. In the analysis presented in this paper, households are defined as separate family groups, with one primary caregiver each. Each of the 229 households was randomly assigned to one of two experimental groups: a treatment group or a control group. In 59 cases, multiple children within a single household were involved in the Family-in-Focus program and received the intervention. However, randomization was clustered at the household level; meaning that within a particular household, all children were assigned to the same intervention group. Randomization was also stratified by household size, the gender and age ratio among eligible children in the sample, and the number of caregivers (i.e. separate households) in the dwelling (Burns, Maughan-Brown, and Mouzinho 2017).<sup>12</sup> During randomization, 106 study households were assigned to the control group, and 123 households were assigned to the treatment group.

**Figure 3: Experimental groups in the HOPE SOAP<sup>®</sup> evaluation**



The evaluation was integrated into the Family-in-Focus program for all 229 households for a period of 12-weeks: from September to December 2014 (Burns, Maughan-Brown, and Mouzinho 2017). During this time, children in both the treatment and control groups received soap and the typical Family-in-Focus programming. Children in treatment group households received HOPE SOAP<sup>®</sup>, whereas children in the control group received a bar of soap almost identical to HOPE SOAP<sup>®</sup> (i.e. transparent, child-sized, colorful), except with a toy alongside, rather than inside, the soap (Burns, Maughan-Brown, and Mouzinho 2017). Ideally, the evaluation would have included a pure control group as well (for example where children in the program did not receive

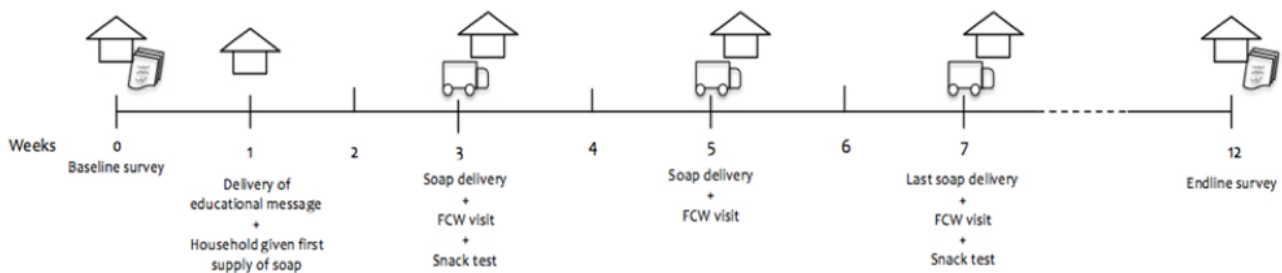
<sup>11</sup> Criterion (2) aims to prevent attrition, and criterion (3) ensures that the effects of HOPE SOAP are not confounded by the effects of another ECD program operating in the same community.

<sup>12</sup> In 22 cases, multiple caregivers reside in the same dwelling. Family groups under separate caregivers which reside in the same dwelling are regarded as separate households.

any soap), but budget constraints and the pilot-nature of the study prevented this. Thus, the results below provide information on the impacts of HOPE SOAP<sup>®</sup> as compared to normal soap.

At the first experimental session in September 2014 (Week 1), children in evaluation households received health and hygiene information from their community health worker along with a bar of soap. After the first session (Week 1), households in both experimental groups continued to receive soap from an independent research team, bi-weekly over an 8-week period (at Weeks 3, 5, 7). Soap delivery coincided with Family-in-Focus visits from the Foundation for Community Work’s health workers. Thus, children received four bars of soap in total throughout the evaluation period (Burns, Maughan-Brown, and Mouzinho 2017). See Figure 4 below for a depiction of the timeline of soap distribution and Family-in-Focus visits.

**Figure 4: Evaluation timeline**<sup>13</sup>



When this experiment was designed by UCT researchers in 2014, its main focus was on children: and the study was therefore aimed at evaluating the direct impact of HOPE SOAP<sup>®</sup> on children’s handwashing behaviors and health outcomes. An initial analysis illustrates that HOPE SOAP<sup>®</sup> does indeed have a positive effect on children’s handwashing behaviors and health (Burns, Maughan-Brown, and Mouzinho 2017). A summary of the impacts of HOPE SOAP<sup>®</sup> on children is presented in a forthcoming section of this paper (Section 3.6). However, as expounded upon previously (Section 2.5), a handwashing program targeted at children may also have indirect impacts on health and behavior which spillover from children to the rest of the household. Accordingly, the handwashing behaviors of the children’s 229 primary caregivers, as well as the health outcomes of the 1,288 non-treated members of children’s households, are of central interest. The analysis presented in this paper explicitly focuses on measuring the indirect effects that a child’s assignment to HOPE SOAP<sup>®</sup> treatment had on the handwashing behaviors and health outcomes of these individuals.

The impacts of HOPE SOAP<sup>®</sup> may spillover to caregivers and other members of the household in the following ways. Firstly, if the health of treated children improves due to HOPE SOAP<sup>®</sup>, all household members

<sup>13</sup> Figure 4 is taken from Burns, Maughan-Brown, and Mouzinho (2017) with permission by the authors.

may benefit from positive health externalities because of the reduction in overall disease transmission rates throughout households (Chung et al. 2015; Angelucci and Di Maro 2010). Second, HOPE SOAP<sup>®</sup> may lead to social interaction spillovers if, for instance, children who receive HOPE SOAP<sup>®</sup> share their soap with family members, indirectly affecting non-treated members' health or handwashing behavior. Finally, behavioral spillovers may occur if HOPE SOAP<sup>®</sup> improves children's handwashing behaviors; thereby inducing other household members to shift their behaviors as well, or changing the overall social norms around handwashing within the treated households (Angelucci and Di Maro 2010).

### **3.3 Data collection**

Data used for the analysis was collected by external enumerators, through questionnaires given to primary caregivers prior to the first soap distribution in September 2014 (Week 0), and again five weeks after the final soap distribution (Week 7) in December 2014 (Week 12) (Burns, Maughan-Brown, and Mouzinho 2017). Refer back to Figure 4 for a depiction of the data collection timeline. In addition to collecting basic demographic information, surveys included detailed questions on handwashing and health. Specifically, questionnaires included indicators of caregivers' self-reported handwashing behaviors, and general health measures for all members of the household. See Table 1 below for details on relevant survey questions and response choices.

Including questions on self-reported handwashing behavior is standard procedure in field research on handwashing with soap (Vindigni, Riley, and Jhung 2011). However, self-reported indicators often provide biased accounts of actual handwashing behavior; and evidence shows that individuals are typically two or three times less likely to wash their hands in practice (based upon observed measures) than what they self-report (Galiani 2012; Orsola-Vidal and Yusuf 2011; Vindigni, Riley, and Jhung 2011). Because of the potential for biased results if only self-reported handwashing behaviors were recorded, this evaluation also included a measure of observed handwashing behavior. At two of the four Family-in-Focus visits which occurred during the experimental period (between Weeks 2-8), community health workers conducted so-called "snack-tests" where they directly observed the handwashing behaviors of children and caregivers in the study. During the snack-tests, health workers presented caregivers and children with a snack of crackers and jam, and recorded whether or not both individuals washed their hands prior to eating (or preparing) the snack (Burns, Maughan-Brown, and Mouzinho 2017).

Health workers were not blinded to treatment status; and it is therefore possible that they could have interacted with households differently and/or recorded the handwashing behaviors (and other observed measures such as caregiver cleanliness) of households differently depending on households' assigned experimental groups. Nevertheless, this evaluation was explicitly designed to minimize any potential bias and its subsequent effects. Firstly, as noted above, soap was distributed to households by an independent research team, rather than the health workers themselves. Second, health workers were under the impression that the goal of the study was to examine the impact of overall soap provision, rather than measure the differences between the impacts of HOPE SOAP<sup>®</sup> and normal soap. And, finally, randomization was stratified at the health worker level, so that each health worker serviced households in both the control and treatment groups.

**Table 1: Survey questions and response choices**

<b>Caregiver Handwashing Behavior</b>			
<i>Variable</i>	<i>Survey Question</i>	<i>Instrument type</i>	<i>Response type</i>
Handwash before food prep	How often do you... Wash hands before preparing food	Self-reported by primary caregiver at baseline and endline	Categorical [Always, Most of the time, Some of the time, None of the time]
Handwash before eating	How often do you... Wash hands before eating	-	-
Handwashing with soap	How often do you use soap for the following activities... To wash your hands	-	-
Handwash at snack-tests	Did ...'s caregiver wash hands before handling food?	Observed by FCW worker at each FIF visit	Dummy variable, equal to 1 if yes
<b>Household Member Health</b>			
<i>Variable</i>	<i>Survey Question</i>	<i>Instrument type</i>	<i>Response type</i>
Health status	How is the health of ... at present? Would you say it is poor, fair, good very good, or excellent?	Reported by primary caregiver at baseline and endline	Categorical [Poor, Fair, Good, Very good, or Excellent]
Diarrhea	In the last 2 weeks, did ... have any diarrhea?	-	Dummy variable, equal to 1 if yes
Flu	In the last 2 weeks, did ... have any flu symptoms (e.g. fever, coughing, sore throat, headache)?	-	-
Nausea	In the last 2 weeks, did...experience any nausea or vomiting?	-	-
Too sick for activities	In the last 2 weeks was ... ever too sick or too ill to carry out his or her normal activities (e.g. play, go to school, work)?	-	-
Number of sick days	How many days in the last 2 weeks was ... too sick or too ill to carry out his or her normal activities?	-	Discrete number

### 3.4 Sample population

Tables 2, 3, and 4 below present baseline demographic information on the sample population. Households in the sample were mainly Coloured and relatively large in size: 82.8% of the heads-of-household were Coloured, and households consisted of 7.7 members on average.<sup>14,15</sup> Additionally, sample households generally had low incomes: over 40% of households reported a total household income of less than R2,000 per month.<sup>16</sup> In 78.5% of cases, households had piped water inside their dwelling; and 87.2% of households reported having a toilet inside their home. Still, 44.4% of households limited their water use, and 18% limited handwashing for members of their households. Significantly more households in the control group had piped water inside their home (84.8% versus 73.2%); and control households were also less likely to limit their water use as compared to treatment households (38.5% versus 50%).<sup>17</sup>

At the time of data collection, survey enumerators also observed whether or not soap was visible in the household, and recorded household cleanliness on a scale between 1 and 10 (where 1 represented a very dirty household and 10 represented a very clean household). These two indicators provide valuable insight into households' general hygiene and handwashing habits; and both cleanliness and soap presence have been used in previous studies as proxies for handwashing behavior when direct observation of handwashing was not possible (Ram 2010; Chase and Do 2010). Enumerators recorded that soap was present in 60.6% of households at baseline. The average cleanliness score of households was a 5.67 out of 10, although households in the treatment group scored slightly higher on the cleanliness measure (6.0) than those in the control (5.2).

Further, over 97% of the primary caregivers in sample households were female, and these caregivers were an average age of 35.7 years old. The highest level of education completed by primary caregivers on average was Grade 10, and only 18.2% of caregivers in the sample reported that they had worked for income in the past two weeks. Moreover, 60.5% of caregivers in the sample had been enrolled in the Foundation for Community Work program for more than 4 months; 62.2% had completed a Family-in-Focus module on health and hygiene in the 3 months prior to baseline data collection; and 35.3% had independently attended a Foundation for Community Work session, in addition to their scheduled Family-in-Focus lessons.

---

<sup>14</sup> “Coloured” is a commonly used racial classification in South Africa, which describes an individual of mixed-race ancestry.

<sup>15</sup> Household size was based upon the number of unique members each caregiver reported living in their household.

<sup>16</sup> Thus, households in the sample are relatively poor. In 2016 average monthly income for Coloured households across South Africa was R14,397 (based on average annual income of R172,765) (Statistics South Africa 2017a).

<sup>17</sup> T-tests are used to detect significant differences between the treatment and control groups.

**Table 2: Household descriptive statistics**

VARIABLE	(1) Full Mean/SD	(2) Control Mean/SD	(3) Treatment Mean/SD
Coloured head of household	0.828 (0.38)	0.827 (0.38)	0.829 (0.38)
Head of household matriculated	0.135 (0.34)	0.133 (0.34)	0.137 (0.35)
Asset score <sup>∇</sup>	9.157 (3.15)	9.186 (3.23)	9.132 (3.10)
Income below 2000	0.407 (0.49)	0.396 (0.49)	0.416 (0.50)
Piped water in home	0.785 (0.41)	0.848 (0.36)	0.732** (0.44)
Toilet in home	0.872 (0.34)	0.904 (0.30)	0.844 (0.36)
Household limits water use	0.444 (0.50)	0.385 (0.49)	0.500* (0.50)
Soap always available	0.717 (0.45)	0.705 (0.46)	0.727 (0.45)
Soap present <sup>†</sup>	0.606 (0.49)	0.660 (0.48)	0.558 (0.50)
Household limits handwashing	0.180 (0.38)	0.190 (0.39)	0.171 (0.38)
Household cleanliness <sup>†</sup> [1-10]	5.677 (2.42)	5.270 (2.46)	6.029** (2.35)
Observations	229	106	123

T-tests for significance: \*p<0.1 \*\*p<0.05 \*\*\*p<0.01

<sup>∇</sup> Composite score of 20 household assets

<sup>†</sup> Variable observed by external enumerators

**Table 3: Caregiver descriptive statistics**

VARIABLE	(1) Full Mean/SD	(2) Control Mean/SD	(3) Treatment Mean/SD
Age	35.780 (11.38)	35.272 (11.20)	36.217 (11.56)
Male	0.027 (0.16)	0.010 (0.10)	0.042 (0.20)
Highest grade completed	10.295 (2.09)	10.277 (2.20)	10.311 (2.00)
Worked for income in past week	0.182 (0.39)	0.149 (0.36)	0.210 (0.41)
Enrolled in FCW for +4 mo.	0.622 (0.49)	0.667 (0.47)	0.583 (0.50)
Received hygiene info in past 3 mo.	0.605 (0.49)	0.653 (0.48)	0.563 (0.50)
Completed a FCW hygiene module	0.353 (0.48)	0.411 (0.49)	0.304 (0.46)
Observations	229	106	123

T-tests for significance: \*p<0.1 \*\*p<0.05 \*\*\*p<0.01

Excluding treated children who received the intervention, 22% of the 1,288 other household members were also children between 2 and 10 years old, and an additional 7% were infants 0-1 years old. Overall the average age for all sample household members (excluding treated children) was 25 years old. Of the adults over 18 years old in the sample, 38% had worked for income in the past week; and 51% of all other individual household members received some type of government grant.<sup>18</sup>

Baseline data collection also demonstrates that random assignment to the treatment and control groups was successful: there were few significant differences between the demographic characteristics of households, caregivers, and individuals in the treatment and control groups. Baseline equivalence confirms that the control group does in fact provide a valid counterfactual for measuring the impact of HOPE SOAP<sup>®</sup> on the treatment group (Gertler et al. 2011; Glennerster and Takavarasha 2013). The few variables which were statistically significantly different between the two groups at baseline will be controlled for during analysis. As noted, these variables included: whether or not a household limited their water use; if a household had piped water in their dwelling; and the observed cleanliness of a household.

<sup>18</sup> The proportion of individuals in the sample who received government grants was higher than the statistic for both the Cape Town Metropolitan area, where only 19% of individuals received some type of government grant in 2016, and South Africa as a whole, where 29.7% individuals received a government grant (Statistics South Africa 2017b).

**Table 4: Individual household member descriptive statistics<sup>19</sup>**

VARIABLE	(1) Full Mean/SD	(2) Control Mean/SD	(3) Treatment Mean/SD
Household size	7.724 (2.84)	7.700 (2.84)	7.745 (2.85)
Age	25.086 (17.82)	25.136 (18.05)	25.043 (17.64)
Male	0.450 (0.50)	0.457 (0.50)	0.444 (0.50)
Highest grade completed, if adult 18+	10.115 (2.78)	10.258 (2.60)	10.000 (2.91)
Worked in past week, if adult 18+	0.389 (0.49)	0.362 (0.48)	0.412 (0.49)
Received social grant	0.511 (0.50)	0.493 (0.50)	0.526 (0.50)
Child age 2-10	0.224 (0.42)	0.237 (0.43)	0.212 (0.41)
Observations	1284	601	683

T-tests for significance: \*p<0.1 \*\*p<0.05 \*\*\*p<0.01

### 3.5 Baseline handwashing and health

Findings from the baseline survey also show that caregivers in the sample population had good handwashing practices and a relatively high knowledge of proper handwashing techniques pre-intervention. These findings are presented in Tables 5 and 6 below. At baseline, 86% of caregivers in the sample stated that they *always* wash their hands before preparing food; 83% stated that they *always* wash their hands before eating; and 88% stated that they *always* use soap to wash their hands. This high level of perfect self-reported behavior is consistent with literature which suggests individuals are likely to report the best possible outcome for handwashing behavior (Vindigni, Riley, and Jhung 2011). Additionally, at baseline, caregivers scored highly on a composite measure of knowledge of proper handwashing techniques and critical times. On a knowledge index between 0 and 11, where 0 indicates no knowledge and 11 indicates perfect knowledge, 52% of caregivers scored a nine or higher; and only 6.5% scored a 4 or lower. Importantly, there were no significant differences in terms of self-reported handwashing behavior or knowledge of proper handwashing techniques between caregivers in the control and treatment groups.

<sup>19</sup> Includes all household members except children ( $n = 288$ ) who were explicitly treated by the intervention.

**Table 5: Caregiver baseline handwashing behaviors**

VARIABLE	(1) Full Mean/SD	(2) Control Mean/SD	(3) Treatment Mean/SD
Wash before food prep [0-3]	2.808 (0.50)	2.802 (0.51)	2.813 (0.50)
Wash before eating [0-3]	2.860 (0.51)	2.886 (0.47)	2.837 (0.55)
Use soap to wash hands [0-2]	1.803 (0.56)	1.840 (0.50)	1.772 (0.61)
Handwashing score [0-6]	5.336 (1.37)	5.413 (1.26)	5.270 (1.46)
Handwashing knowledge [0-11]	8.483 (2.48)	8.574 (2.50)	8.407 (2.47)
Observations	229	106	123

T-tests for significance: \* $p < 0.1$  \*\* $p < 0.05$  \*\*\* $p < 0.01$

Moreover, at baseline, caregivers reported the health status and prevalence of illness symptoms for all members of their households. The baseline health outcomes of all 1,288 non-treated members of the household, including caregivers themselves but excluding treated children ( $n = 288$ ), are shown in Table 6. Overall, caregivers reported that household members were generally in good health. On a scale of Poor, Fair, Good, Very Good, or Excellent, caregivers reported that 49% of household members' health was Good. Only 14% of individuals' health was Poor or Fair; and, 36.5% of individuals' health was Very Good or Excellent (see Figure 5).<sup>20</sup> However, household members in the treatment group were reported to have significantly better health than those in the control group ( $p$ -value of 0.07). Thus, this difference will also be controlled for during endline analysis.

However, even when individuals were reported to have less than Good health at baseline, rarely did caregivers report that an individual's health affected their ability to participate in their normal day-to-day activities. Only 6.5% of household members in the sample were too sick to take part in their normal day-to-day activities in the two weeks prior to the baseline survey. Of that 6.5%, the average number of days missed due

<sup>20</sup>Although caregivers reported that most individuals did have at least Good health at baseline, individuals in the sample actually have slightly worse reported health than the average Coloured South African. According to the 2016 Demographic Health Survey, 31% of all Coloured South Africans are in Good health, and 59.7% report Very Good or Excellent health (Statistics South Africa 2017b).

to poor health was 5.8 days. Additionally, fewer than 3% of individuals in the sample population experienced nausea or diarrhea symptoms in the two weeks prior to baseline. Instances of flu were more prevalent for individuals in the sample: around 16% of household members experienced some flu symptom (such as cough or fever) in the previous two weeks. Thus, at the individual level, the sample population was generally healthy. However, 53% of individuals belonged to a household where some household-member was ill in the previous two weeks; and 9% of individuals belonged to a household where a child under-10 who was not enrolled in the Family-in-Focus program was ill in the previous two weeks.

**Table 6: Baseline individual household member health**

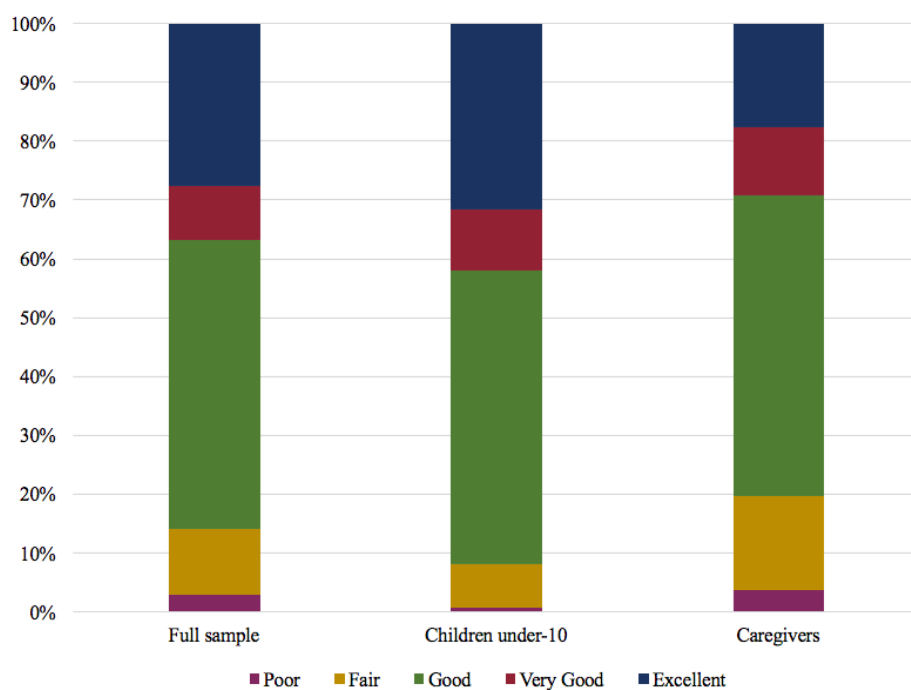
VARIABLE	(1) Full Mean/SD	(2) Control Mean/SD	(3) Treatment Mean/SD
Health status [1-5]	3.475 (1.09)	3.415 (1.09)	3.528* (1.09)
Too sick for day-to-day	0.065 (0.25)	0.066 (0.25)	0.064 (0.25)
Number of sick days	0.353 (1.78)	0.367 (1.87)	0.341 (1.71)
Diarrhea	0.021 (0.14)	0.024 (0.15)	0.018 (0.13)
Flu	0.157 (0.36)	0.163 (0.37)	0.151 (0.36)
Nausea	0.029 (0.17)	0.035 (0.18)	0.025 (0.15)
Other child in household has illness	0.092 (0.29)	0.097 (0.30)	0.087 (0.28)
Other household member has illness	0.537 (0.50)	0.535 (0.50)	0.539 (0.50)
Observations	1284	601	683

T-tests for significance: \*p<0.1 \*\*p<0.05 \*\*\*p<0.01

Despite the fact that sample population seemed relatively healthy at baseline, it should be noted that because caregivers reported on the health of all household members, the indicators used to assess health may be relatively blunt measures of an individual's actual health. It is probable that caregivers were more likely to know about their own health and the health of their young children, than the health of other adult members in

their household. This could be one reason why caregivers had slightly worse reported-health than the full sample population (see Figure 5); or why there were more reported symptoms of flu than reported cases of diarrhea (assuming caregivers were more likely to know about another’s flu symptoms than diarrhea instance).

**Figure 5: Individuals’ reported health status at baseline**



### 3.6 Child level results

As previously mentioned, the results and discussion presented in this paper examine the effects that a child’s assignment to HOPE SOAP<sup>®</sup> treatment has on caregivers’ handwashing behaviors and household members’ health. However, when considering the forthcoming results and their implications, it is important to keep in mind how HOPE SOAP<sup>®</sup> affected treated children in the study. Burns, Maughan-Brown, and Mouzinho (2017) presented the results from an initial analysis of the child level data in their SALDRU working paper entitled *“Washing with Hope: Evidence from a hand-washing pilot study among children in South Africa.”*<sup>21</sup> Findings published in the working paper affirm that the intervention did, in fact, have a positive effect on children’s handwashing behaviors and health outcomes. Children in the both the treatment (HOPE SOAP<sup>®</sup>) and control (normal soap) groups experienced improvements in handwashing behavior from baseline to endline. Caregivers from both experimental groups reported that children who received the soap were more likely to wash their hands before a meal and after using the toilet, and were more likely to use soap when washing their hands, post-intervention. Additionally, conditional on poor baseline handwashing behavior, children who received HOPE SOAP<sup>®</sup> were significantly more likely to wash their hands after using the toilet than children who received

<sup>21</sup> SALDRU is the Southern African Labor & Development Unit housed within the School of Economics at UCT.

normal soap. HOPE SOAP<sup>®</sup> children also scored significantly higher on an index of reported soap use at endline than children in the control group (Burns, Maughan-Brown, and Mouzinho 2017).

Moreover, children in the HOPE SOAP<sup>®</sup> treatment group were more likely to be observed washing their hands unprompted before eating a snack as compared to control children, though these results were not statistically significant. Treatment children were 6 percentage points more likely to wash their hands at a second snack-test (p-value 0.48), and 10 percentage points more likely to wash their hands at both snack-tests (p-value 0.26) than children who received normal soap (Burns, Maughan-Brown, and Mouzinho 2017). Evidence also showed that all children in the sample were healthier at endline than at baseline. And, furthermore, children who received HOPE SOAP<sup>®</sup> displayed fewer symptoms of illness at endline than children who received normal soap. On a composite index of illness symptoms, which ran from 0 to 13, HOPE SOAP<sup>®</sup> children displayed - 0.35 fewer symptoms of illness than children in the control group (p-value 0.25). Although many of these results lack statistical significance, the findings reported by Burns, Maughan-Brown, and Mouzinho (2017) are certainly promising.

## 4. ANALYSIS AND RESULTS

### 4.1 Estimation strategy

To identify the indirect effects that HOPE SOAP<sup>®</sup> treatment has on the handwashing behaviors and health of non-treated household members, this paper applies the following Ordinary Least Squares (OLS) regression model to each outcome of interest:

#### Equation 1:

$$y_i = \alpha + \beta T + \delta X + \varepsilon_i$$

In Equation 1, the dependent variable  $y_i$  is the outcome of interest (either health or behavior) for each individual caregiver or household member  $i$ ;  $\delta$  is a vector of coefficients on a group of control variables  $X$ ;  $\alpha$  is some constant; and  $\varepsilon_i$  is the unobserved error term for each individual observation  $i$ . Covariates in  $X$  include the demographic characteristics observed to be unbalanced between the control and treatment groups at baseline. When caregiver behaviors are the dependent variables,  $X$  includes: (1) whether or not a household has piped water inside their dwelling; (2) whether or not a household limits water use; and (3) a household's baseline cleanliness score. When  $y_i$  is an individual household member's health outcome,  $X$  also includes (4) an individual's baseline health status. The parameter of interest is  $\beta$ : the coefficient on the dummy independent variable  $T$ ; where  $T$  is equal to 1 if individual  $i$  comes from a household where a child was assigned to the treatment group, and equal to 0 if individual  $i$  comes from a household assigned to the control group. The size and sign of  $\beta$  represent the effect that assignment to HOPE SOAP<sup>®</sup> treatment has on the outcome of interest,  $y$ , for each individual,  $i$ . Recall that, in this experiment, children in households assigned to the control group

received normal soap, and thus  $\beta$  should be interpreted as the effect of HOPE SOAP<sup>®</sup> relative to normal soap. All standard errors presented in the following results are clustered at the community health worker level in order to account for potential error correlation between households served by the same health worker. Clustering the standard errors at the health worker level also corrects for heteroskedasticity in the data, or non-constant error variance (Deaton 1997).<sup>22</sup>

In many cases in the analysis below, the outcomes of interest,  $y_i$ , are binary variables (i.e. equal to either 0 or 1). For example, observed handwashing behavior is equal to 1 if a caregiver washed their hands at the snack-test, and equal to 0 if they did not. When  $y_i$  is a binary variable, Equation 1 above should be interpreted as a linear probability model (LPM). In the LPM,  $\beta$  estimates the change in the probability that  $y_i=1$  resulting from a change in assignment to the treatment group. Therefore, in the case of observed handwashing behavior,  $\beta$  estimates the change in the probability that a caregiver will wash their hands due to treatment assignment.

Further, because many of the outcomes of interest are binary, and can only be equal to 0 or 1, they are inherently non-linear. Linear regression models, such as OLS or LPM, are meant to estimate linear and continuous variables; and therefore, some statistical problems may arise when applying linear estimation techniques to binary dependent variables. For instance, linear regression models may give predicted values of  $y_i$  that fall outside the range of possible outcomes [0,1];  $\beta$  coefficients may grossly underestimate (or overestimate) true effect sizes; and statistical corrections that traditionally would improve the quality of the estimates (i.e. reporting results with robust standard errors) may actually make linear model estimators less efficient (Angrist and Pischke 2009; Wooldridge 2009). Accordingly, this paper also employs logistic (logit) and probability unit (probit) regression analysis to models with binary dependent variables (described below and adapted from Wooldridge 2010).

Nevertheless, although estimating the effect of a treatment on a non-linear variable using OLS/LPM may cause some statistical problems, linear regression models typically still provide good approximations of the partial effects of a treatment on binary dependent variables (Wooldridge 2010). Furthermore, OLS/LPM regression models provide results which are much simpler to interpret than binary choice models, and can be more easily understood by policy makers and practitioners alike (Wooldridge 2009). Accordingly, this paper primarily uses the  $\beta$  coefficients from the OLS/LPM models to interpret the effects of treatment assignment; and relies on logit and probit models to ensure the robustness of the OLS/LPM results.

A logit regression model uses a logarithm as a link function to predict the probability of achieving an outcome of interest,  $y_i$  (i.e. the probability that  $y_i=1$ ). As shown in Equation 2 below, the  $\beta$  coefficient in the logit model is the maximum likelihood estimator associated with the probability of achieving the outcome of interest  $y_i$ , conditional on the treatment,  $P(y_i=1|T)$ :

---

<sup>22</sup> OLS estimates of  $\beta$  remain unbiased and consistent in the presence of heteroskedasticity, though they are inefficient (Deaton 1997; Wooldridge 2009).

**Equation 2:**

$$P(y_i=1|T) = \frac{e^{\alpha + \beta + \delta X}}{1 + e^{\alpha + \beta + \delta X}}$$

As the function above is non-linear, the  $\beta$  coefficient itself does not represent the effect of the treatment on the outcome of interest. Rather, the Equation 3 must be used to reveal the average marginal, or partial, effect of the treatment,  $T$ , on the outcome of interest,  $y_i$ :

**Equation 3:**

$$APE = \frac{e^{\alpha + \beta + \delta X}}{1 + e^{\alpha + \beta + \delta X}} - \frac{e^{\alpha + \delta X}}{1 + e^{\alpha + \delta X}}$$

Similar to how a logit uses a logarithm function to predict the probability that an outcome of interest  $y_i = 1$ , a probit regression model uses a cumulative normal distribution,  $\Phi$ , to produce a probability for  $y_i$  between 0 and 1; where given any Z-score,  $\Phi(Z) \in [0,1]$ . The normal distribution,  $\Phi$ , is then used as a link function and substituted into the basic OLS model to solve for the probability that  $y_i = 1$ :

**Equation 4:**

$$y_i = \Phi(\alpha + \beta T + \delta X + \varepsilon_c)$$
$$\Phi^{-1}(y_i) = \alpha + \beta T + \delta X + \varepsilon_c$$

Hence,  $\beta$  is representative of the change in the Z-score of  $y_i$  when the  $T$  changes from 0 to 1 (or a shift from the control to treatment group). In the probit model, the marginal effect of the treatment on  $y_i$  can be calculated using the link function  $\Phi$ , where the probability of the outcome of interest occurring, that  $y_i=1$ , is equal to:

**Equation 5:**

$$P(y_i=1|T) = \Phi\left(\frac{-\beta' T}{\sigma}\right)$$

And the average marginal effect of the treatment is equal to:

**Equation 6:**

$$APE = \beta \Phi(\alpha + \beta T + \delta X + \varepsilon_c)$$

Thus, when the effects of the treatment on binary dependent variables are interpreted below, the results presented will include the treatment coefficients from all three regression models (OLS/LPM, Logit, Probit), and the accompanying average partial effect sizes given by the binary choice models.

## 4.2 Self-reported versus observed handwashing behavior

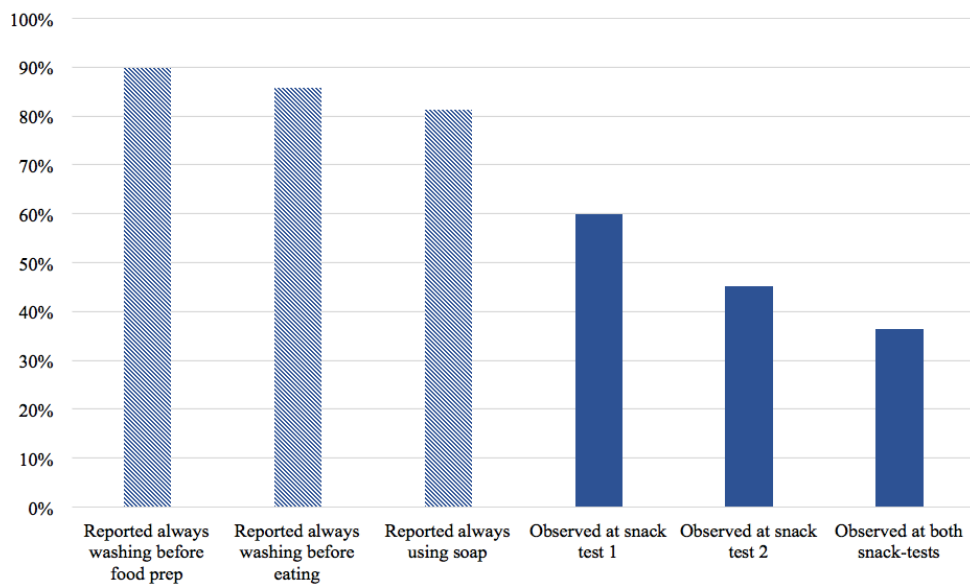
Although a large amount of data exists on caregivers' self-reported handwashing behaviors, a careful investigation of the data showed that self-reported behaviors may provide poor approximations of caregivers' actual behaviors. At endline, 82% of caregivers reported that they both *always* wash their hands before preparing food and *always* use soap to wash their hands. However, in practice, only 36.4% of caregivers were observed by the community health workers to wash their hands before both snack-tests. This is a 45.6 percentage point discrepancy. Figure 6 below shows the inconsistencies between the proportion of caregivers who claimed perfect behavior in the self-reported measures, and the percent of caregivers who were actually observed to wash their hands at the snack-tests.

These large discrepancies could exist for a few reasons. First, the difference between caregivers' self-reported and observed behaviors is consistent with literature which claims that self-reported measures of handwashing typically grossly overstate actual handwashing behavior (Galiani, Gertler, Orsola-Vidal 2012; Orsola-Vidal and Yusuf 2011; Vindigni, Riley, and Jhung 2011). Interestingly, in the experimental sample, the correlation coefficient between a composite index of self-reported handwashing behavior and actual instance of observed handwashing at both snack tests is -0.18.<sup>23</sup> This negative correlation coefficient suggests that caregivers who reported good handwashing behaviors were actually slightly less likely to wash their hands at both snack-tests in actuality. Therefore, it is quite probable that the self-reported handwashing measures are biased upwards. Second, the large differences between self-reported and observed behaviors could be due to the design of the snack-test itself. It is possible that caregivers do actually have high rates of handwashing behavior (as they self-report), but the snack-tests conducted during the experiment simply did not provide an accurate representation of their true behaviors. For instance, perhaps caregivers do typically wash their hands before preparing food, but in the case of the snack-test, they did not physically handle the snack and were therefore not observed to wash their hands. Despite this possibility, random assignment (and the fact that each health worker serviced both treatment and control households) ensures that the potential measurement error due to the design of the snack-test should be equivalent across both experimental groups. Thus, comparing the observed handwashing behaviors of caregivers in the treatment and control groups will still provide an accurate depiction of the effect of the treatment. Accordingly, results presented in this paper will not rely on any self-reported measures of handwashing, but rather will solely use observed data from the two snack-tests to accurately assess the impact of HOPE SOAP<sup>®</sup> on caregivers' handwashing behaviors.

---

<sup>23</sup> A correlation coefficient of 1 (-1) indicates perfect positive (negative) correlation between two variables, while a correlation coefficient of 0 indicates no correlation.

**Figure 6: Reported handwashing behavior vs. observed instance of handwashing**



#### **4.3 Result 1: HOPE SOAP<sup>®</sup> improves caregiver handwashing behavior at snack-test 2**

The means of caregivers' observed handwashing behaviors at the snack-tests are presented in Table 7 below. As illustrated in Table 7, there was no significant difference in the proportion of caregivers who washed their hands at the first snack-test between the treatment and control groups. However, by snack-test 2, 51.8% of caregivers in the treatment group were observed washing their hands, but only 36.9% of caregivers in the control group engaged in the behavior. A rank-sum test confirms that this difference was statistically significant: caregivers whose children received HOPE SOAP<sup>®</sup> washed their hands significantly more at snack-test 2 than caregivers whose children receive control soap (p-value 0.07).<sup>24</sup>

Furthermore, a simple comparison of means also shows that the percentage of caregivers who washed their hands declined over time (from snack-test 1 to snack-test 2) for caregivers in both experimental groups. However, the rate of decline was much greater for caregivers in the control group than for those in the treatment. In the control group, the number of caregivers who washed their hands decreased by 27 percentage points from snack-test 1 to snack-test 2. Conversely, instance of handwashing only declined by 5.1 percentage points over time for caregivers in the treatment group (see Figure 7). This trend is similar to the findings for children's handwashing behavior. Burns, Maughan-Brown, and Mouzinho (2017) reported that the number of handwashing-children in the control group decreased by 8 percentage points from snack-test 1 to snack-test 2 (from 56% to 48%), while the number of handwashing-children in the treatment group did not decline at all (59%).

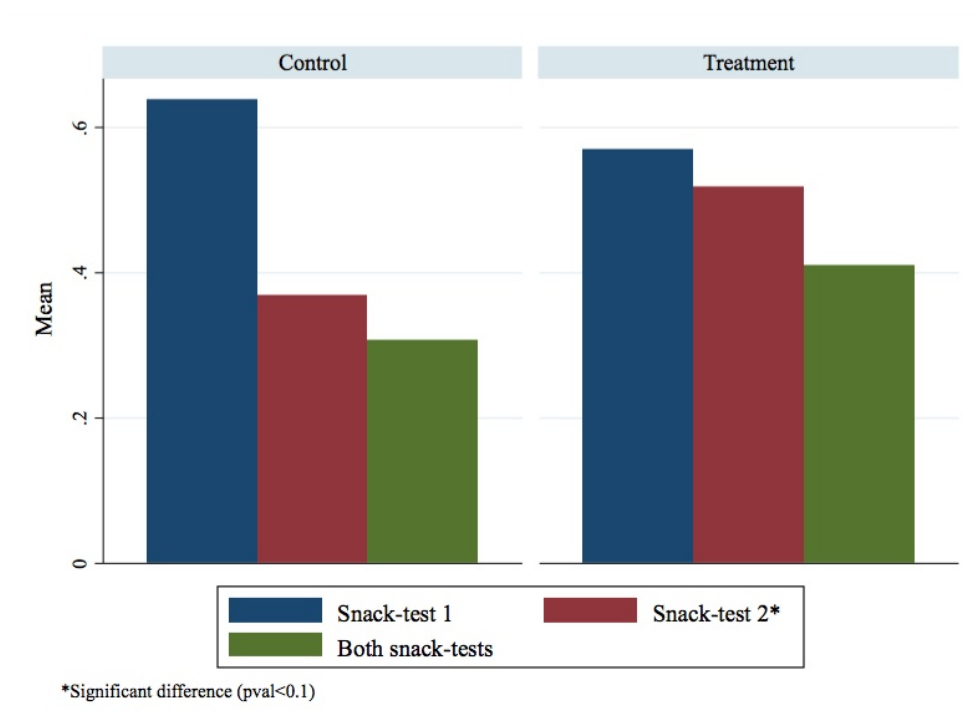
<sup>24</sup> A rank-sum test for significance is used rather than a traditional t-test because it does not assume that the variable of interest is normally distributed (binary variables cannot be normally distributed as they can only equal values [0,1]).

**Table 7: Means of caregiver handwashing behavior<sup>25</sup>**

VARIABLE	(1) Control Mean/SD	(2) Treatment Mean/SD	(3) Glass's D Diff./CI-95%
Snack-test 1	0.639 (0.48)	0.569 (0.50)	0.144 [-0.14, 0.43]
Snack-test 2	0.369 (0.49)	0.518* (0.50)	-0.306 [-0.63, 0.02]
Both snack-tests	0.308 (0.47)	0.410 (0.49)	-0.219 [-0.54, 0.10]
Observations	83	102	229

\*p<0.1 \*\*p<0.05 \*\*\*p<0.01

**Figure 7: Observed handwashing at snack-tests**



<sup>25</sup> In an effort to present results which are comparable across studies, standardized effect sizes (the standard difference between the two groups' means) are displayed in column (3) and calculated using Glass's Delta:  $\frac{\mu_{treatment} - \mu_{control}}{\sigma_{control}}$

OLS regression analysis, completed using the model presented in Equation 1 above, also demonstrates that HOPE SOAP<sup>®</sup> had a positive effect on caregiver behavior at snack-test 2. Results from the regression analysis are shown in Table 8 below. The evidence suggests that holding all else equal, assignment to HOPE SOAP<sup>®</sup> treatment leads to a 13 percentage point increase, on average, in the probability that a caregiver will wash their hands at snack-test 2 (column 2). This result is only marginally significant with a p-value of 0.17. However, in the case of this pilot evaluation, a lack of statistical significance using the standard critical value threshold of 0.10 does not necessarily mean that assignment to HOPE SOAP<sup>®</sup> will not positively affect caregiver behavior. One downfall of testing interventions using randomized experiments with small sample sizes is the fact that small samples are often underpowered to pick up the true effects of the treatment when they do exist (Gertler et al. 2011). This consideration, in addition to the fact that 0.13 is a relatively large effect size, suggests that despite having no statistical significance, it is quite possible that HOPE SOAP<sup>®</sup> could have a sizable positive effect on caregiver behavior in practice.

**Table 8: Effect of HOPE SOAP<sup>®</sup> on caregiver handwashing behavior (OLS results)**

VARIABLES	(1) Snack-test 1	(2) Snack-test 2	(3) Snack-test 2	(4) Both tests
Treatment	-0.10 (0.08) 0.23	0.13 (0.09) 0.17	0.14* (0.08) 0.08	0.07 (0.08) 0.42
Snack-test 1			0.43*** (0.08) 0.00	
Piped water in house	0.08 (0.11) 0.44	-0.08 (0.11) 0.46	-0.10 (0.09) 0.24	-0.04 (0.10) 0.74
Baseline cleanliness score	0.07*** (0.01) 0.00	0.07*** (0.02) 0.00	0.03 (0.02) 0.14	0.08*** (0.02) 0.00
Limits water use	0.07 (0.08) 0.37	0.09 (0.09) 0.33	0.08 (0.08) 0.30	0.12 (0.09) 0.19
Constant	0.18 (0.13) 0.17	0.02 (0.13) 0.87	-0.00 (0.12) 0.97	-0.14 (0.11) 0.21
Observations	146	119	119	119
R-squared	0.13	0.14	0.29	0.19

Standard errors in parentheses, followed by p-values

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

The treatment coefficients from regression models which examine caregiver handwashing behavior at snack-test 1 and at both snack-tests (columns 1 and 4) are also statistically insignificant, and have larger p-values than the coefficient on the treatment from snack-test 2. Nevertheless, these models suggest that a child's assignment to treatment leads to a 7 percentage point increase in the probability of handwashing at both snack-tests (p-value 0.42); but a 10 percentage point decrease in the probability of handwashing at snack-test 1 (p-value 0.23). The negative coefficient for the model with snack-test 1 as the dependent variable could indicate that HOPE SOAP<sup>®</sup> may have had an adverse effect on behavior at the first snack-test, although this negative effect was reversed by the second snack-test. Alternatively, as another possibility, because snack-test 1 was the first directly observed measure of handwashing behavior, it is also plausible that the negative coefficient is due to baseline differences in actual handwashing behavior (as opposed to self-reported behavior) between caregivers in the control group and caregivers in the treatment group. There is no way to know for certain if this was the case because observed handwashing behavior was not recorded at baseline. However, the comparison of means table (Table 6) shows that the mean of handwashing behavior at snack-test 1 was higher for caregivers in the control group (63.9%) than for caregivers in the treatment group (56.9%).

Furthermore, as past actions are a key determinant of current habitual behaviors, it also is reasonable to assume that a caregiver's behavior at snack-test 2 is likely a function of their behavior at snack-test 1 (Ouellette and Wood 1998; Neal et al. 2012; Aunger et al. 2010). For this reason, and because observed handwashing behavior at snack-test 1 may serve as a proxy for baseline actual handwashing, this paper also runs an OLS regression model that includes the variable for handwashing behavior at snack-test 1 as an additional independent variable in the regression of the treatment on behavior at snack-test 2. Output from this model is also presented in Table 8 above (column 3). When added to the specification, the coefficient on caregiver behavior at snack-test 1 is statistically significant at the 1% level: suggesting that a caregiver's behavior at snack-test 2 is determined, in part, by their behavior at snack-test 1. Holding all else equal (including treatment status), if a caregiver washes their hands at snack-test 1, the probability that they will also wash their hands at snack-test 2 is 43 percentage points higher than if they had not washed their hands at snack-test 1. Moreover, the output of this model also shows that when caregiver behavior at snack-test 1 is held constant, assignment to HOPE SOAP<sup>®</sup> treatment leads to a significant 14 percentage point increase, on average, in the probability that a caregiver will wash their hands at snack-test 2 (p-value 0.08). Hence, this model reaffirms inference made from the basic regression model: that assignment to HOPE SOAP<sup>®</sup> treatment improves caregiver handwashing behavior at snack-test 2.

Table 9 reports the treatment coefficients and average partial effect sizes derived from logistic and probability unit models which regress the treatment on observed caregiver behavior variables (as described in Equations 2 through 6). As shown in Table 9, in all cases, the average partial effects of the treatment from the binary models are consistent with the  $\beta$  coefficients from the linear regression models (OLS/LPM). That is, that the average marginal effects of the treatment from the logit and probit models are of the same sign, and around a similar size, to the treatment coefficients given by the OLS/LPM regressions. Thus, the OLS/LPM results are robust to a number of specifications.

**Table 9: Effect of HOPE SOAP<sup>®</sup> on caregiver handwashing behavior (OLS/Logit/Probit results)**

VARIABLES	(1) OLS	(2) Logit	(3) Logit APE	(4) Probit	(5) Probit APE
<b>Snack-test 1</b>					
Treatment coefficient	-0.10	-0.50	-0.10	-0.31	-0.10
<i>Standard error</i>	(0.08)	(0.41)	(0.08)	(0.25)	(0.08)
<i>P-value</i>	0.23	0.23	0.22	0.21	0.20
R <sup>2</sup> / Hosmer-Lemeshow pval	0.13	0.23		0.25	
Observations	146	146	146	146	146
Controls	Yes	Yes	Yes	Yes	Yes
<b>Snack-test 2</b>					
Treatment	0.13	0.58	0.12	0.37	0.12
	(0.09)	(0.42)	(0.08)	(0.26)	(0.08)
	0.17	0.17	0.16	0.15	0.15
R <sup>2</sup> / Hosmer-Lemeshow pval	0.14	0.20		0.68	
Observations	119	119	119	119	119
Controls	Yes	Yes	Yes	Yes	Yes
<b>Snack-test 2</b>					
Treatment	0.14*	0.78*	0.13*	0.46*	0.14*
	(0.08)	(0.44)	(0.07)	(0.26)	(0.07)
	0.08	0.08	0.07	0.08	0.07
Snack-test 1	0.43***	2.13***	0.37***	1.26***	0.38***
	(0.08)	(0.48)	(0.05)	(0.28)	(0.06)
	0.00	0.00	0.00	0.00	0.00
R <sup>2</sup> / Hosmer-Lemeshow pval	0.29	0.75		0.87	
Observations	119	119	119	119	119
Controls	Yes	Yes	Yes	Yes	Yes
<b>Both snack-tests</b>					
Treatment	0.07	0.34	0.06	0.23	0.07
	(0.08)	(0.43)	(0.08)	(0.26)	(0.08)
	0.42	0.43	0.43	0.38	0.37
R <sup>2</sup> / Hosmer-Lemeshow pval	0.19	0.04		0.04	
Observations	119	119	119	119	119
Controls	Yes	Yes	Yes	Yes	Yes

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

#### **4.4 Result 2: HOPE SOAP<sup>®</sup> may improve caregiver handwashing behavior through behavioral spillovers from treated children**

As extrapolated upon previously, HOPE SOAP<sup>®</sup> treatment may lead to changes in caregiver behavior through spillovers which occur when caregivers are indirectly treated by the intervention soap, despite it being targeted specifically at children. Firstly, interaction effects could have occurred if treated children shared soap with their caregivers, thereby also inducing caregivers to wash their hands. However, only 2% of caregivers in the evaluation sample reported that their child shared his/her soap with anyone else in the household. Therefore, it can be assumed that interaction spillovers did not occur in the context of this evaluation. Second, theoretically, HOPE SOAP<sup>®</sup> could cause behavioral spillovers in handwashing. Recall that behavioral spillovers are defined as indirect effects on a non-treated population that arise either when an intervention (1) changes the behaviors of individuals in close proximity to one another, or (2) changes the social norms surrounding a particular behavior within a treated context (Angelucci and Di Maro 2010). Accordingly, by improving the handwashing practices of children in the treatment group, HOPE SOAP<sup>®</sup> could implicitly encourage caregivers to wash their hands as well, or shift overall household social norms towards better handwashing practices.

There is evidence that the handwashing behaviors of individuals in the evaluation sample were determined, in-part, by household-level behaviors: caregivers' and children's handwashing behaviors at each snack-test were highly correlated. As shown in Table 10 below, at snack-test 1 the correlation coefficient between child and caregiver observed handwashing behavior for the full sample population was 0.80. While the correlation coefficient was close to 0.80 for both treatment and control households at snack-test 1, at snack-test 2 the correlation between caregiver and child behavior was much stronger for individuals in the treatment group than in the control group. At snack-test 2 (both snack-tests), the correlation between the behaviors of individuals in the treatment group increased to 0.90 (0.82), while it fell to 0.62 (0.56) for individuals in the control group. This may be evidence that HOPE SOAP<sup>®</sup> generated a stronger social norm around handwashing behavior in treated households, as compared to households who received normal soap.<sup>26</sup>

---

<sup>26</sup> It would be valuable to be able to quantify HOPE SOAP<sup>®</sup>'s effect on household social norms, although this is difficult to do in practice. One way researchers have attempted to measure handwashing norms is through examining changes in rates of self-reported handwashing (Hussam et al. 2016; Caroli and Weber-Baghdiguian 2016). Since self-reported measures can serve as a proxy for the social desirability of the practice, changes in rates of self-reported handwashing can provide insight into the social norms around handwashing behavior (Hussam et al. 2016). In this evaluation however, there are very little differences in self-reported handwashing measures from baseline to endline, and no significant differences between the self-reported rates for the treatment and control groups.

**Table 10: Correlation coefficients between child and caregiver behavior**

	Caregiver	Child		
		Snack-test 1	Snack-test 2	Both
(1) Full sample <i>n = 145</i>	Snack-test 1	0.802		
	Snack-test 2		0.785	
	Both			0.718
(2) Treatment <i>n = 82</i>	Snack-test 1	0.800		
	Snack-test 2		0.903	
	Both			0.829
(3) Control <i>n = 62</i>	Snack-test 1	0.813		
	Snack-test 2		0.626	
	Both			0.562

Whether or not HOPE SOAP<sup>®</sup> may affect caregiver behavior by way of a behavioral spillover from treated children can be explored by considering children’s behavior as a mediator for caregivers’ behavior – or a mechanism by which the treatment indirectly influences the outcome. This relationship can be assessed empirically, using a mediation analysis technique to examine the effect of child behavior on caregiver behavior. If child behavior were a mediator for caregiver behavior, statistically the following three criteria would be true (Baron and Kenny 1986; Kazdin 2007). First, HOPE SOAP<sup>®</sup> would influence caregiver behavior. Result 1 above provides evidence of this at snack-test 2: assignment to HOPE SOAP<sup>®</sup> improves caregiver behavior at snack-test 2. Second HOPE SOAP<sup>®</sup> would influence child behavior. The results reported by Burns, Maughan-Brown, and Mouzinho (2017) suggest that HOPE SOAP<sup>®</sup> did have positive impacts on children’s handwashing behaviors, justifying this criterion. And finally, third, when added to the regression of HOPE SOAP<sup>®</sup> on caregiver behavior, (a) the coefficient on child behavior would be statistically significant, and (b) including the child level variable would reduce the effect of the treatment on caregiver behavior. In other words, in Equation 7 below, where  $M$  represents the mediating variable of child behavior,  $\lambda$  would be statistically significant and  $\beta_m$  would be smaller than  $\beta$  from Equation 1.

**Equation 7:**

$$y_i = \alpha + \beta_m T + \lambda M + \delta X + \varepsilon_i$$

Table 11 below shows the output from the mediation analysis (Equation 7) which regresses the treatment and child behavior on caregiver handwashing behavior at each snack-test. Indeed, for the outcomes of snack-test 2 and both snack-tests, (columns 4, 5, 7) the coefficients on child behavior are statistically significant at the 1% level (satisfying criterion 3a). Furthermore, for snack-test 2 and both snack-tests, the

coefficients on the treatment are smaller than those from the original OLS models (column 3 and 6) (satisfying criterion 3b).<sup>27</sup> At snack-test 2, the coefficient on the treatment declines from 0.13 to 0.05; and for both snack-tests the treatment coefficient declines from 0.07 to 0.02. Thus, children’s behavior meets all the necessary statistical conditions to be considered a mediator for caregiver behavior at snack-test 2 and both snack-tests; and it is possible that child behavior mediates the effect of the treatment on caregiver behavior.<sup>28</sup>

Despite this apparent causal chain, inferences made based on the results of the mediation analysis should be handled with caution. First, because data were collected on caregiver and child behaviors at the same point in time, there is no way to discern if child behavior changed prior to caregiver behavior and not vice versa (Kazdin 2007). Second, although results from Burns, Maughan-Brown, and Mouzinho (2017) certainly suggest that HOPE SOAP<sup>®</sup> led to improvements in children’s handwashing behavior, the small sample size and lack of statistical significance in results means that this cannot be unequivocally be proven true. Nevertheless, the mediation analysis, in conjunction with the stronger correlation coefficients present between caregiver and child behavior in the treatment group at snack-test 2 and at both snack-tests, suggests that HOPE SOAP<sup>®</sup> may indirectly affect caregiver behavior through behavioral spillovers from treated children.

**Table 11: Child behavior as a mediator for HOPE SOAP<sup>®</sup>’s effect on caregiver behavior (OLS results)**

VARIABLES	(1) Snack-test 1		(3)	(4) Snack-test 2		(5)	(6) Both snack-tests	
	(2)	(7)						
Treatment	-0.10 (0.08)	-0.15*** (0.05)	0.13 (0.09)	0.05 (0.05)	0.05 (0.05)	0.07 (0.08)	0.02 (0.07)	
	0.23	0.01	0.17	0.32	0.32	0.42	0.77	
Child wash at ST-1		0.75*** (0.06)						
		0.00						
Child wash at ST-2				0.75*** (0.05)	0.77*** (0.06)			
				0.00	0.00			
Snack-test 1					-0.02 (0.05)			
					0.65			
Child wash at both							0.65*** (0.07)	
							0.00	
Observations	146	143	119	118	118	118	116	
R-squared	0.13	0.61	0.14	0.65	0.65	0.19	0.55	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	

Standard errors in parentheses, followed by p-values. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

<sup>27</sup> As considered in Result 1, it is not clear what the outcome variable of wash at snack-test 1 captures; and further, in Table 10, the correlation coefficients between child and caregiver behavior are similar at snack-test 1, suggesting that spillovers may not yet have occurred at snack-test 1. Accordingly, no further discussion of the results in column 2 is warranted.

<sup>28</sup> These results are robust to the output from logit and probit binary regression models (see Appendix 1).

#### 4.5 **Result 3: HOPE SOAP<sup>®</sup> has differential effects on behavior, conditional on baseline cleanliness**

Recall that baseline household cleanliness is included as a covariate in the models above because the cleanliness scores of caregivers in the treatment and control groups were not equivalent at baseline. Further, in many of the regression outputs presented above, the coefficients on baseline household cleanliness are statistically significant at the 1% level (e.g. Table 8, columns 1, 2, 4). Thus, baseline household cleanliness is a significant determinant of caregiver handwashing behavior. Additionally, upon further exploration, it is apparent that baseline cleanliness is positively correlated with observed handwashing behavior. The correlation coefficients between cleanliness and observed caregiver behavior were: 0.4, 0.34, and 0.38 for behavior at snack-test 1, snack-test 2, and both snack-tests respectively. Based on a review of the literature, this positive relationship should be expected. For example, Ram (2010) noted that observations of a household's environment were a valuable way to gather information about a household's typical handwashing and hygiene behaviors; and Chase and Do (2010) used caregiver cleanliness as a proxy for handwashing behavior when they were unable to observe caregivers' behaviors in practice.

Accordingly, as a household's level of cleanliness can provide insight into their overall handwashing and hygiene habits, and also because there is a positive correlation between caregiver behavior and baseline observed cleanliness, it is possible that HOPE SOAP<sup>®</sup> may have differential effects on caregiver behavior depending on household cleanliness. In the evaluation of HOPE SOAP<sup>®</sup>, household cleanliness was recorded by survey enumerators on a scale from 1-10. At baseline data collection, the variable for cleanliness had a mean of 5.67 and a median of 5. Therefore, households who scored 6 or higher on the cleanliness measure at baseline can be considered relatively clean, and those who scored 5 or lower can be classified as relatively un-clean. To test for the possibility of heterogeneous treatment effects by baseline household cleanliness, the following model, which includes an interaction term, is estimated:

##### **Equation 8:**

$$y_i = \alpha + \beta_1 T + \beta_2 Z + \beta_3 TZ + \delta X + \varepsilon_c$$

In Equation 8,  $Z$  is a cleanliness dummy variable equal to 0 if a household's baseline cleanliness was less than or equal to 5, and 1 if baseline cleanliness was equal to 6 or higher. The interaction term  $TZ$ , represents the cleanliness dummy  $Z$  multiplied by the treatment dummy  $T$ . In Equation 8, a statistically significant  $\beta_3$  coefficient on  $TZ$  indicates that the effect of the treatment  $T$  differs across groups (Brambor, Clark, and Golder 2006). To correctly interpret the effect of HOPE SOAP<sup>®</sup> using Equation 8, the marginal effect of the treatment must subsequently be calculated using Equation 9 below. In Equation 9, the effect of the treatment,  $T$ , on outcome  $y_i$  is equal to  $(\beta_1 + \beta_3)$  when condition  $Z$  is met; and  $\beta_1$  when condition  $Z$  is not met (Brambor, Clark, and Golder 2006).

##### **Equation 9:**

$$\frac{\partial y}{\partial T} = \beta_1 + \beta_3 Z$$

Although a model which includes interaction terms is the most appropriate way to detect heterogeneous treatment effects, the power to pick up statistical significance where it does exist decreases drastically when interaction terms are introduced into a model (Brookes et al. 2004). As this experiment's power to detect statistical significance is already quite low due to its small sample size, it is worth noting that a model which includes interaction terms is likely underpowered to detect any differential treatment effects that do occur.

Outputs from the interaction models are shown in Table 12 below. All coefficients on the interaction terms are negative, suggesting that the effect of the treatment on caregiver behavior is less when households are cleaner at baseline, than when households are dirtier at baseline (Brambor, Clark, and Golder 2006). This may be due, in part, to the fact that caregivers whose households are cleaner have better overall hygiene and are more likely to wash their hands; and therefore, there may be less space for the treatment to have a positive effect on behavior when households are clean at baseline.

Moreover, the marginal effect of the treatment on caregiver behavior at snack-test 2 is -0.02 when the condition of cleanliness is met, and 0.19 when the condition of cleanliness is not met (column 4). This discrepancy is quite sizable (0.21), and implies that HOPE SOAP<sup>®</sup> has a large differential effect on caregiver behavior at snack-test 2, conditional on baseline household cleanliness. Although the marginal effect of the treatment on clean-household caregivers is negative (-0.02), because it is of a small effect size, and because caregivers who are cleaner at baseline are also more likely to wash their hands in general, it should not be too concerning. The 0.19 marginal effect of the treatment on caregivers whose households are less clean at baseline, on the other hand, is quite substantial. The 0.19 marginal effect of the treatment suggests that, conditional on poor household cleanliness at baseline (i.e. poor existing hygiene habits), a child receiving HOPE SOAP<sup>®</sup> leads to a 19 percentage point increase in the probability of a caregiver handwashing at snack-test 2.<sup>29</sup>

---

<sup>29</sup> The coefficients produced by the interaction model are robust to a specification which uses the basic OLS model (Equation 1), but solely examines the effect of the treatment on the subpopulation of households ( $n=138$ ) who are unclean at baseline ( $\beta_T = 0.16$ ,  $SE = 0.11$ ,  $pval = 0.14$ ); and robust to output from binary regression models. See Appendix 2 and 3 for tables detailing the results.

**Table 12: Differential effects of HOPE SOAP<sup>®</sup> by baseline household cleanliness (OLS results)**

VARIABLE	(1) (2)		(3) (4) (5)			(6) (7)	
	Snack-test 1		Snack-test 2			Both snack-tests	
Treatment	-0.10 (0.08) 0.23	0.01 (0.12) 0.92	0.13 (0.09) 0.17	0.19 (0.12) 0.12	0.15 (0.10) 0.15	0.07 (0.08) 0.42	0.16 (0.10) 0.12
Clean household		0.35*** (0.10) 0.00		0.35*** (0.12) 0.00	0.19 (0.13) 0.14		0.36*** (0.11) 0.00
Clean*Treatment		-0.23 (0.15) 0.12		-0.21 (0.17) 0.20	-0.11 (0.16) 0.49		-0.22 (0.15) 0.15
Snack-test 1					0.38*** (0.08) 0.00		
Baseline cleanliness	0.07*** (0.01) 0.00		0.07*** (0.02) 0.00			0.08*** (0.02) 0.00	
Observations	146	171	119	139	139	119	139
R-squared	0.13	0.09	0.14	0.10	0.23	0.19	0.13
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, followed by pvalues

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

#### 4.6 **Result 4: HOPE SOAP<sup>®</sup> improves caregiver responsiveness to the FCW program**

In addition to observing handwashing behavior at the snack-tests, at each of the four Family-in-Focus visits during the intervention period, health workers also recorded caregivers' levels of interest in the week's lesson on a scale of 1 to 10. Further, at endline, caregivers reported the number of other Foundation for Community Work sessions they attended, aside from their scheduled Family-in-Focus visits, throughout the 12-week study period.<sup>30</sup> These program-involvement indicators may provide valuable insight into how caregivers respond to the Foundation for Community Work program as a whole. Table 13 below shows that caregivers whose children received HOPE SOAP<sup>®</sup> had a greater level of interest in the four Family-in-Focus lessons on average, as compared to control caregivers (although this difference was not statistically significant). Additionally, at endline, caregivers in the treatment group were also more likely to have attended a Foundation for Community Work session, separate from their scheduled Family-in-Focus visits, during the 12-week intervention period; and treatment group caregivers attended more sessions overall than those in the control group. As health workers serviced households in both the treatment and control groups, these differences are unlikely to have arisen because of differences in health worker disposition, or simply due to bias in health worker reporting.

<sup>30</sup> A table detailing the indicator and response choices is given in Appendix 4.

**Table 13: Means of caregiver responsiveness variables**

VARIABLE	(1)	(2)	(3)
	Control Mean/SD/N	Treatment Mean/SD/N	T-test Diff/Pval/N
Average interest at all visits [1-10]	7.16 (0.31) 55	7.74 (0.22) 71	-0.57 (0.12) 126
Ever attended a FCW session throughout the study period	0.75 (0.82) 65	0.92 (0.80) 81	-0.17 (0.20) 146
Number of FCW sessions attended	0.64 (0.09) 65	0.77 (0.08) 81	-0.13 (0.32) 146
Interest at first FIF visit [1-10]	7.67 (0.30) 81	7.90 (0.22) 98	-0.22 (0.53) 179

Moreover, Table 14 contains the output from OLS regression analyses which estimate the effects of the treatment on the program-involvement indicators. As shown in Table 14 (column 1), holding all else equal, assignment to HOPE SOAP<sup>®</sup> treatment leads to a significant increase in caregiver interest in the Family-in-Focus program by 0.88 units on average (p-value 0.02). Because the baseline survey did not include an initial measure of caregiver’s interest, to ensure this regression output does not only capture baseline differences in interest between caregivers in the control and treatment groups, an additional regression is presented in column 2, which uses data on caregiver interest at the first of the four Family-in-Focus visits as a pseudo-baseline measure. A t-test for balance on this pseudo-baseline measure (see Table 13) affirms that the interest of caregivers in the treatment and control groups was not statistically different at the first checklist visit. Nevertheless, column 2 displays the results of a model which includes interest at visit-one as a covariate in the regression of the treatment on interest at visits 2, 3, and 4. When the variable which may account for potential baseline differences is included in the specification, the effect of the treatment remains positive: again suggesting that HOPE SOAP<sup>®</sup> improves caregiver interest in the Family-in-Focus program as a whole.<sup>31</sup>

Furthermore, HOPE SOAP<sup>®</sup> treatment also increases the probability that a caregiver will attend a Foundation for Community Work session by 18 percentage points on average (column 4, p-value 0.24),<sup>32</sup> and leads caregivers in the treatment group to attend 0.15 more sessions than those in the control group (column 5,

<sup>31</sup> Note that the size of the treatment effect is smaller (and insignificant) in this specification. One reason for this may be because the treatment already has an effect on interest at visit-one, and thus the covariate for interest at visit-one captures part of the treatment effect. Nevertheless, including the pseudo-baseline measure helps to rule out the possibility that the regression output only captures baseline differences in interest – strengthening the validity of the result from column 1.

<sup>32</sup> The treatment coefficient on the regression of the binary variable of whether or not a caregiver had ever attended a FCW session is robust to the APEs from both logit and probit regression models (for both models:  $APE = 0.16$ ,  $SE = 0.08$ ,  $pval = 0.06$ ).

p-value 0.31).<sup>33</sup> Although it is difficult to tease out the reasons for these results due to a lack of qualitative data on the matter, the fact that HOPE SOAP<sup>®</sup> increases caregiver interest and attendance in the Foundation for Community Work program is particularly interesting seeing as caregivers in both groups received the exact same Family-in-Focus curriculum, and children in both groups received the same toys and soap (shape and size) – with the only difference being whether or not the child was given HOPE SOAP<sup>®</sup> or soap with a toy alongside of it.

**Table 14: Effect of HOPE SOAP<sup>®</sup> on caregiver responsiveness (OLS results)**

VARIABLES	(1) Average interest	(2) Avg interest at visits 2, 3, 4	(4) Ever attend FCW session	(5) Num sessions attended
Treatment	0.88** (0.37)	0.28 (0.41)	0.18 (0.16)	0.15 (0.16)
	0.02	0.49	0.24	0.31
Interest at FIF visit-one		0.37*** (0.11)		
		0.00		
Observations	100	94	116	116
R-squared	0.19	0.29	0.11	0.29
Controls	Yes	Yes	Yes	Yes

Standard errors in parentheses, followed by pvalues

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

#### 4.7 **Result 5: The treatment’s effect on behaviors is not mediated by their interest or knowledge**

Because HOPE SOAP<sup>®</sup> leads caregivers in the treatment group to have a greater interest in the Family-in-Focus program than those in the control group, it may be that interest is a latent mechanism through which the intervention indirectly influences caregiver behavior (Baron and Kenny 1986; Kazdin 2007). To empirically examine if a caregiver’s level of interest influences their handwashing behavior, interest can be included as an additional independent variable in regressions of the treatment on behavior (following the same mediation analysis specification presented in Section 4.4, Equation 7). Output from these additional regressions, shown in Table 15 below, illustrate that this is not the case: caregiver interest does not exert influence on caregiver behavior at the snack-tests, nor mediate HOPE SOAP<sup>®</sup>’s effect on behavior (columns 1, 3, 5, 7). Coefficients on the interest variables are small in size, indicating that interest has little effect on caregiver behavior; and the treatment coefficients in the new models are relatively the same size as they are in the models where interest is

<sup>33</sup> No additional covariates were included in the regressions of the treatment on caregiver attendance of FCW sessions because the three variables measuring program attendance (i.e. enrolled in FCW for +4 mo., received hygiene info in past 3 mo., and completed a FCW hygiene module) were all equivalent at baseline (see Table 3).

excluded.<sup>34</sup> Thus, it does not appear that a caregiver’s handwashing behavior is influenced by their interest in the Family-in-Focus program as a whole.

The same is true for caregivers’ knowledge about proper handwashing techniques and critical times. Caregivers’ knowledge about handwashing did rise slightly from baseline to endline for caregivers in both experimental groups: by 5% on average (on a composite index of 0-11).<sup>35</sup> However, when added to the regression specification, knowledge does not influence the treatment’s effect on caregiver behavior (columns 2, 4, 6, 8). This result reaffirms the idea present throughout literature: that knowledge (and/or beliefs and intentions) does not influence an individual’s actions when behaviors are habitual. Thus, in contrast Result 2 which suggests that HOPE SOAP<sup>®</sup> may indirectly affect caregiver behavior through behavioral spillovers from treated children, the results presented in Table 15 suggest that assignment to HOPE SOAP<sup>®</sup> treatment has an independent effect on caregiver behavior which is not mediated by improvements in knowledge of proper handwashing techniques, nor by interest in the Foundation for Community Work program as a whole.<sup>36</sup>

**Table 15: Interest and knowledge additional covariates in the regressions on behavior (OLS results)**

VARIABLES	Snack-test 1		Snack-test 2				Both snack-tests	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.11 (0.09)	-0.09 (0.08)	0.14 (0.11)	0.13 (0.09)	0.21** (0.09)	0.14* (0.08)	0.10 (0.10)	0.07 (0.09)
	0.23	0.27	0.19	0.17	0.02	0.08	0.32	0.42
Interest	0.01 (0.02)		0.00 (0.02)		-0.00 (0.02)		-0.01 (0.02)	
	0.64		0.95		0.85		0.63	
Knowledge		0.02 (0.02)		0.00 (0.02)		-0.01 (0.02)		0.02 (0.02)
		0.28		0.90		0.48		0.41
Snack-test 1					0.55*** (0.09)	0.44*** (0.08)		
					0.00	0.00		
Observations	99	145	95	119	89	119	95	119
R-squared	0.24	0.14	0.09	0.14	0.30	0.29	0.16	0.19
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, followed by pvalues. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

<sup>34</sup> Furthermore, in Table 15 - columns 5, when snack-test 1 is included as an additional covariate in the model, adding FIF interest to the regression of the treatment on behavior at snack-test 2 actually strengthens the positive impact of HOPE SOAP on caregiver handwashing behavior.

<sup>35</sup> There were no significant differences between the baseline or endline knowledge scores for caregivers in the treatment versus control groups.

<sup>36</sup> These results are robust to the output from binary regression models (see Appendix 5 and 6).

#### 4.8 **Result 6: A child's assignment to HOPE SOAP<sup>®</sup> has no effect on household member health**

In addition to the treatment indirectly affecting caregiver handwashing behavior, household member health may be impacted by a child's assignment to HOPE SOAP<sup>®</sup>. Recall that spillovers from treated children to other household members may occur as a result of social interactions or externalities. Firstly, interaction effects could lead to improved health for household members if a treated child shared his/her soap with other individuals, thereby also providing them with increased access to soap. However, recall that only 2% of caregivers reported that their child shared soap with anyone else in the household. Thus, in practice it is highly unlikely that interaction spillover effects impacted household member health. Secondly, externalities, or positive health effects that occur when disease transmission is lowered in a treated context, could lead to positive spillovers for individual household-member health (Angelucci and Di Maro 2010). In their 2017 working paper, Burns, Maughan-Brown, and Mouzinho (2017) found that HOPE SOAP<sup>®</sup> improved the general health of treated children. Therefore, it is possible that individuals residing in HOPE SOAP<sup>®</sup> households may have also experienced positive health impacts as a result of reduced disease transmission.

The linear regression model given in Equation 1 is used to estimate the impacts of HOPE SOAP<sup>®</sup> on the health outcomes of all additional members of the household (including sample caregivers, but not treated children). Despite the theoretical possibility that HOPE SOAP<sup>®</sup> could influence the health of all members of the households, output from the OLS/LPM models, provided in Table 16 below, demonstrates that a child's assignment to HOPE SOAP<sup>®</sup> does not lead to any particularly large, or statistically significant, effects on household member health. Additionally, the signs of the treatment coefficients are inconsistent across the scope of all health outcomes, and thus no inference can be made as to whether or not HOPE SOAP<sup>®</sup> has an overall positive or negative effect on individual health.

It is also worth noting that the R-squareds for all the regression models which examine HOPE SOAP<sup>®</sup>'s impact on health are very low (between 0.00 and 0.07). This means that only between 0% and 7% of the variation in health outcomes is able to be attributed to the models' specifications. The models may not be able to accurately predict individual health for a variety of reasons. For one, it is possible that treatment assignment and handwashing behavior are not important determinants of individual health within the evaluation context (i.e. other environmental factors not included in the specifications are more important predictors of health). Further, recall that health outcomes were reported by primary caregivers at endline data collection, and as such, may be imprecise measures of individual health. Accordingly, even if the R-squareds were higher, they might still give biased estimates of actual individual health. Section 5.3 in the discussion below further investigates reasons why the regression models may be unable to provide insights on how HOPE SOAP<sup>®</sup> affects individual health. Nevertheless, the treatment coefficients presented in Table 16 are consistent with the outputs from binary regression models (see Appendix 7). That is, the treatment coefficients from the linear models are similar to the average partial effects produced by logit and probit regression models.<sup>37</sup>

---

<sup>37</sup> Hence, it is unlikely that the linear nature of the OLS/LPM model (versus a binary model) is a reason for the low R<sup>2</sup>'s.

**Table 16: Effect of HOPE SOAP<sup>®</sup> on household member health (OLS results)**

VARIABLES	(1) Health status	(2) Diarrhea	(3) Flu	(4) Nausea	(5) Any symptom	(6) Too sick for actv	(7) Num sick days
Treatment	0.03 (0.13) 0.82	-0.00 (0.01) 0.84	0.04 (0.03) 0.18	-0.01 (0.01) 0.29	0.03 (0.04) 0.38	-0.02 (0.02) 0.30	-0.09 (0.13) 0.47
Piped water in house	0.23 (0.17) 0.17	-0.00 (0.02) 0.94	-0.09* (0.05) 0.10	-0.01 (0.02) 0.52	-0.06 (0.06) 0.25	0.01 (0.03) 0.85	0.16 (0.12) 0.16
Baseline cleanliness	-0.07*** (0.03) 0.01	-0.00 (0.00) 0.93	0.00 (0.01) 0.71	0.00 (0.00) 0.58	0.00 (0.01) 0.96	0.00 (0.01) 0.44	-0.02 (0.03) 0.54
Limits water use	0.02 (0.13) 0.90	0.00 (0.02) 0.77	0.03 (0.03) 0.33	-0.00 (0.01) 0.82	0.02 (0.04) 0.66	0.00 (0.02) 0.94	0.01 (0.12) 0.93
Baseline health	-0.22*** (0.05) 0.00	-0.00 (0.01) 0.95	-0.01 (0.01) 0.33	-0.01 (0.01) 0.32	-0.01 (0.01) 0.31	-0.02** (0.01) 0.02	-0.15** (0.06) 0.02
Constant	3.55*** (0.28) 0.00	0.04 (0.02) 0.12	0.16*** (0.06) 0.01	0.05* (0.02) 0.07	0.20*** (0.07) 0.00	0.15*** (0.05) 0.01	0.82*** (0.27) 0.00
Observations	772	773	773	773	774	774	772
R-squared	0.07	0.00	0.02	0.00	0.01	0.01	0.01

Standard errors in parentheses, followed by pvalues. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Furthermore, although it does not have an effect on household level health, because HOPE SOAP<sup>®</sup> leads to positive impacts on caregiver handwashing behavior at snack-test 2, it is conceivable that the intervention could have had a positive effect on caregiver health. However, the total number of caregivers in the sample who reported symptoms of illness was very low. For instance, only 8 caregivers reported having nausea in the two weeks prior to endline data collection (3 in the control group, and 5 in the treatment group). The small absolute number of caregivers who reported symptoms of illness means that the data lacks variation; and therefore, regression analysis will not be able to provide any robust inferences as to the effects of the treatment. Additionally, in the sample, caregiver handwashing behavior was not correlated with better reported health, but rather the opposite. The correlation coefficients between caregiver handwashing and health actually suggest that handwashing was correlated with more reported symptoms of illness.<sup>38</sup> This again alludes to the

<sup>38</sup> The correlation coefficients between caregiver handwashing behavior at snack-test 2 and health outcomes are all positive (Flu = 0.05, Nausea = 0.08, Diarrhea = 0.10). Recall that health outcomes were recorded as binary variables, equal to 1 if the caregiver reported experiencing a symptom of illness.

possibility that handwashing may not be a significant determinant of health in the context of this evaluation; and coupled with the lack of variation in the data, suggests that any analysis done on caregivers' health would not produce results with sufficient explanatory power.

#### **4.9 Summary of results**

In sum, results from the empirical analysis conducted above conclude that a child's assignment to HOPE SOAP<sup>®</sup> treatment has a positive impact on the handwashing behaviors of their caregiver, but no effect on the health outcomes of other household members. Assignment to HOPE SOAP<sup>®</sup> treatment improves caregiver handwashing behavior at snack-test 2 by approximately 13 percentage points. This positive effect is not affected by caregivers' knowledge of proper handwashing techniques, nor their interest in the Family-in-Focus program as a whole. However, evidence indicates that HOPE SOAP<sup>®</sup> may strengthen handwashing norms within treated households; and, that the effect of HOPE SOAP<sup>®</sup> on caregivers' handwashing behaviors may be mediated by children's behaviors. Moreover, results suggest that HOPE SOAP<sup>®</sup> has a larger effect on caregivers who have poor hygiene habits at baseline: conditional on poor baseline cleanliness, the treatment leads to a 19 percentage point increase in the probability that a caregiver will wash their hands at snack-test 2. Further, caregivers whose children receive HOPE SOAP<sup>®</sup> respond significantly better to the Foundation for Community Work program as a whole than those whose children receive normal soap. Nonetheless, despite the positive effect that HOPE SOAP<sup>®</sup> has on caregiver handwashing behavior, the intervention has little effect on the health outcomes of individuals in the treated households.

## **5. DISCUSSION**

### **5.1 Limitations and caveats**

The randomized design of this impact evaluation, in addition to many of the statistical precautions taken during analysis, ensure both experimental and analytical rigor, as well as robust results. However, several caveats and limitations should be noted. First, the pilot nature of the evaluation limited both the number of distinct intervention arms possible in the design, and the total sample size. As previously mentioned, because of its relatively small sample size, this evaluation may be underpowered to pick up statistically significant results, even when treatment effects do exist. Moreover, a 13% attrition rate further reduced the sample's size; although 13% attrition is relatively standard throughout the impact evaluation literature.<sup>39</sup> In total, 31 of the 229 households enrolled in the evaluation were absent at endline. Nevertheless, there is no evidence of differential, or non-random, attrition between households in the treatment and control groups.

---

<sup>39</sup> For instance, Cameron, Shah, and Olivia (2013) have an attrition rate of 8.5%; Galiani, Gertler, and Orsola-Vidal (2012) experience 20% attrition; and experiments conducted by Banerjee et al. (2007) have 17 and 18% attrition.

Because of low power, it is difficult to discern if many of the statistically insignificant behavioral findings presented above exist because of a true lack of effect, or because of the small sample size (Haushofer and Shapiro 2016). If pilot constraints were not an issue, researchers would have ideally conducted power calculations before beginning the experiment to determine the sample size necessary to pick up significant effects on handwashing and health. Still, completing power calculations ex-post can provide a basis for considering whether or not null results occur because of a true lack of effect, or due to insufficient power (Haushofer and Shapiro 2016). One component of a power calculation is the minimum detectable effect (MDE) size, which provides researchers with the sample size necessary to pick up statistically significant results of a particular minimum size (Duflo, Glennerster, and Kremer 2006; Haushofer and Shapiro 2016). Calculating MDEs ex-post, using the actual sample size to compute the MDE necessary for obtaining significance, provides a benchmark for interpreting null results (Haushofer and Shapiro 2016). The ex-post MDE for caregiver behavior at snack-test 2 is 0.22. Thus, in order to pick up statistical significance with the existing sample size, the treatment must lead to a 22 percentage point change in the probability that a caregiver will wash their hands at snack-test 2. Similarly, by substituting the actual observed effect size,  $\hat{\beta}$ , into the MDE equation, the sample size which would have been necessary to obtain statistical significance at  $\hat{\beta}$  can be computed. For an effect of 0.13 to be statistically significant (with p-value < 0.10), the study would have had to evaluate the effects of HOPE SOAP<sup>®</sup> on 352 caregivers.<sup>40</sup> Because the 0.13 effect size delineated in Result 1 is much smaller than the MDE of 0.22, is still marginally significant (and significant when the covariate of snack-test 1 is included in the regression model), and is observed with a smaller sample than the 352 observations which would have been necessary for significance, it can be inferred that the evaluation is simply underpowered to pick up the positive impact of the treatment on caregiver behavior.

In addition to limiting sample size, the evaluation's pilot nature restricted the number of intervention arms in the experimental design. As such, the evaluation contained no pure control group, and inference can only be made on the effect of HOPE SOAP<sup>®</sup> as compared to a group which also received soap and Family-in-Focus programming. Accordingly, the results provide little insight into the effects of the intervention as a whole (soap + program enrollment); and no explicit conclusions can be made on how HOPE SOAP<sup>®</sup> might impact members of a population who are not also receiving some sort of early childhood development programming which includes a health and hygiene module. While this aspect of the design is a limitation in some respects, it also gives this study a distinct advantage over impact evaluations which examine the effects of handwashing package interventions (including soap, education, etc.). Studies which examine the effects of package interventions may be unable distinguish between the effects of each individual intervention component. Therefore, as previously discussed (in Section 2), evaluations of comprehensive interventions are less useful

---

<sup>40</sup> MDEs were calculated using the following equation adapted from Haushofer and Shapiro (2016) and Duflo, Glennerster, and Kremer (2006):  $MDE = (t_{1-\kappa} + t_{\alpha/2}) \frac{\sigma}{\sqrt{NP(1-P)}}$ . Using the standard significance levels and powers of 0.10 and 0.80 respectively, the equation simplifies to:  $MDE = 2.48 \times SE(\hat{\beta})$ . When the MDE equation is used to calculate ex-post necessary sample sizes,  $\sigma$  is the observed standard error, and P is equal to the proportion of the population that was treated. Tables which include ex-post MDE's and sample sizes for each snack-test outcome can be found in Appendix 8.

from a policy perspective, as providing entire package interventions at scale may be too expensive and too logistically challenging for developing countries to implement (Zwane and Kremer 2007). In this way, the design of this evaluation contributes to the existing literature by explicitly testing HOPE SOAP<sup>®</sup> (which could be more easily delivered in practice) against an alternative soap product.

Lastly, the pilot nature of the study also limited the time period in which the experiment occurred to three months. Handwashing behavior was observed within a 7-week time period, and health outcomes were observed at Week-12 (5 weeks after the last soap distribution). Recall from the handwashing literature that behavior change programs are often successful in creating short-term positive impacts on behavior, but not long-term improvements. In the long-run, handwashing behaviors often regress back to pre-intervention levels, partly due to the habitual nature of handwashing, and the fact that individuals often relapse into past habits over time (Neal et al. 2015; Vindigni, Riley, and Jhung 2011; Chase and Do 2010). This means that the handwashing outcomes observed within the 7-week time period of this evaluation may not be a true representation of effects of the program on handwashing with soap in the long-run. Future research should attempt to include longer-term follow-ups on handwashing behavior.

In addition to the limitations caused by the experiment's design, a few additional statistical caveats should be mentioned. First, as previously discussed (Result 6), the goodness-of-fit measures (R-squareds and Hosmer-Lemeshow tests) for the models which regress the treatment on health outcomes are extremely low: meaning that the models used do not explain the data very well. Section 5.3 below investigates the implications of poor goodness-of-fit in depth. Finally, it is worth mentioning the possibility of family-wise error, also known as the multiple-outcome problem. As the outcomes of interest considered in each section of the analysis belong to the same family (i.e. all behavioral outcomes or all health outcomes), testing them simultaneously expands the probability of falsely rejecting a null hypothesis and claiming significance when it does not exist (Duflo, Glennerster, and Kremer 2006; Holm 1979; Lee and Shaikh 2014). As the number of regressions performed on outcomes within the same family increases, so does the probability of falsely inferring significance.<sup>41</sup> Further, including numerous iterations of regression models on the same dependent variable may be considered specification searching. Specification searching threatens the legitimacy of a study when researchers run multiple models with the hopes of finding statistical significance in one to report. (Christensen and Miguel 2016). Despite this potential issue, one of the main goals of any pilot evaluation is to thoroughly explore all possible research questions in an attempt to understand how a treatment affects a sample population; and then to use this deep understanding as a basis for conducting an evaluation at scale in the future. Thus, because of the pilot nature of this evaluation, a thorough exploration of the data is warranted. Running numerous regression specifications helps to uncover the mechanisms by which HOPE SOAP<sup>®</sup> affects health and behavioral outcomes. Gaining a deeper understanding of these mechanisms will ultimately lead to the design of a better

---

<sup>41</sup> One possible solution is to apply the Bonferroni correction, and lower the critical value to  $\alpha/n$ , where  $n$  represents the number of tests conducted. However, the Bonferroni correction comes at the expense of greatly increasing the possibility of encountering a Type II error, or failing to find significant effects when they do exist (Gelman, Hill, and Yajima 2012).

intervention and experiment in the future. Accordingly, the discussion of results below details lessons learned from the evaluation, which can be used as a basis for future work. Furthermore, this paper reports null findings alongside significant ones, and thus attempts to be as transparent about the effects of HOPE SOAP<sup>®</sup> as a possible. Keeping the aforementioned statistical and experimental caveats in mind, the remainder of Section 5 presents a discussion of the results delineated in Section 4.

## **5.2 The promise of habitual behavior change**

The results presented above show that assignment to HOPE SOAP<sup>®</sup> treatment leads to an increase in the probability that a caregiver washes their hands at snack-test 2 by 13 percentage points on average (Result 1). This finding is promising, and in conjunction with evidence from the rest of the analysis on caregiver behavior, provides reason to believe that HOPE SOAP<sup>®</sup> causes habitual handwashing behavior change which could persist in the long-run. The evidence is favorable for a variety of reasons. First, not only does HOPE SOAP<sup>®</sup> lead to a higher probability of engaging in handwashing behavior at snack-test 2, but it leads to a significantly higher probability when controlling for behavior at snack-test 1 (Result 1). Recall from the literature on habit (Section 2.4) that past actions are a key determinant in predicting current habitual behaviors (Ouellette and Wood 1998). While this is certainly the case for caregivers in the sample (caregivers who wash at snack-test 1 are 43 percentage points more likely to wash at snack-test 2), HOPE SOAP<sup>®</sup> is also successful in inducing handwashing when controlling for a caregiver's prior-behaviors. Furthermore, HOPE SOAP<sup>®</sup> has a larger impact on caregiver behavior when caregivers' households are unclean at baseline (Result 3). Assuming that a dirty household represents poor hygiene habits more generally, this again alludes to the fact that HOPE SOAP<sup>®</sup> may be able to disrupt caregivers' existing poor-hygiene habits. Literature suggests that the ability of an intervention to disrupt existing habits is key to creating lasting change in habitual behaviors (Neal et al. 2015). Hence, the fact that HOPE SOAP<sup>®</sup> has a stronger effect on caregivers whose households were dirty at baseline, in addition to the finding that HOPE SOAP<sup>®</sup> is able to significantly improve caregiver behavior at snack-test 2 when controlling for behavior at snack-test 1, is certainly encouraging.

Furthermore, one reason interventions are often ineffective at changing habitual behaviors is because they concentrate on updating an individual's beliefs and intentions, rather than attempting to leverage automatic processes to disrupt old habits (Marteau, Hollands, and Fletcher 2012). Updating beliefs and intentions does not lead to behavior change in habits because habitual behaviors are not driven by cognitive thought. This is why programs which provide education on the benefits of handwashing with soap often fail to create lasting behavior change, even when they are successful at improving knowledge (Marteau, Hollands, and Fletcher 2012; Curtis, Danquah, and Aunger 2009). Hence, throughout the literature, knowledge is rarely a predictor of actual handwashing behavior (Biran et al. 2009; Vindigni, Riley, and Jhung 2011). This is also true in the context of HOPE SOAP<sup>®</sup>: caregivers' behaviors are not determined by their knowledge of proper handwashing practices; nor does knowledge affect the degree of influence that the treatment has on behavior (Result 5). Therefore, it can be inferred that HOPE SOAP<sup>®</sup> does not change handwashing behavior through the cognitive process of updating caregivers' knowledge or beliefs.

Because it does not influence caregiver behavior by improving knowledge, HOPE SOAP<sup>®</sup> must affect handwashing behavior through some alternative causal pathway. One way the analysis above explores this is by interrogating the relationship between caregiver and child behavior (Result 2). The positive and strong correlation between caregiver and child handwashing behavior suggests that handwashing behavior is consistent throughout households. As such, handwashing behavior in the evaluation households may be considered a social norm. Recall that social norms often serve as the contextual cues which induce habitual behaviors (Curtis, Danquah, and Aunger 2009). Thus, since handwashing is a habitual behavior, partially determined by social norms, a positive correlation between child and caregiver behavior should be expected. However, while the correlation between behaviors at snack-test 1 is strong for the entire sample, the correlation between child and caregiver behavior at snack-test 2 (and at both snack-tests) is much stronger in households where children receive HOPE SOAP<sup>®</sup>. Hence, it is possible that one way HOPE SOAP<sup>®</sup> improves caregiver behavior at snack-test 2 is by strengthening the social norms surrounding handwashing within treated households. As discussed, interventions which can alter the underlying processes which lead to habitual behaviors are typically more successful in creating lasting change (Neal et al. 2015; Marteau, Hollands, and Fletcher 2012). Thus, because social norms are an underlying driver of habits, and HOPE SOAP<sup>®</sup> strengthens the social norms within treated households, HOPE SOAP<sup>®</sup> shows promise of being able to generate sustained handwashing behavior change.

Furthermore, as HOPE SOAP<sup>®</sup> was specifically designed to target children, its impacts on caregiver behavior can be considered a behavioral spillover, or an effect that occurs when an intervention changes the social norms within a treated context (Angelucci and Di Maro 2010). Findings from the mediation analysis provide reason to believe that child behavior could be one mechanism through which HOPE SOAP<sup>®</sup> influences caregiver behavior (Result 2).<sup>42</sup> This evidence, of behavioral spillovers, is important from a research perspective for two reasons. First, it shows that positive spillovers exist in handwashing interventions; and thus, studies which fail to account for spillovers in design and analysis could underestimate the true impacts of handwashing encouragement programs. Second, because there are currently no studies which discuss spillovers from handwashing promotion programs in the development and health economics literature, the evaluation of HOPE SOAP<sup>®</sup> makes a significant contribution. From a policy perspective, the fact that HOPE SOAP<sup>®</sup> may influence handwashing behavior through behavioral spillovers is also imperative. If the effects of a handwashing encouragement intervention can transfer throughout a population, policy-makers can design better, and more cost-effective, programs with this in mind. Nevertheless, because of the caveats explained in Section 4 above (Result 2), more research is needed to prove that behavioral spillovers do exist as a consequence of the provision of HOPE SOAP<sup>®</sup> before policy makers should apply these lessons in practice.

Moreover, recall that in order to create lasting behavior change, handwashing interventions need to (1) disrupt existing poor-handwashing behaviors and, equally as important, (2) ensure that new behaviors stick

---

<sup>42</sup> However, as previously noted, data on the relationship between caregiver and child behavior lacks a clear timeline which would unequivocally prove that child behavior change occurs first, and subsequently causes caregiver behavior change. This limitation serves as a lesson for future research into behavior change: researchers should design instruments for data collection in a way that allows analysis to disentangle the causal chain of events.

(Neal et al. 2015). Thus, the fact that HOPE SOAP<sup>®</sup> causes handwashing behavior to be upheld from snack-test 1 to snack-test 2 for caregivers in the treatment group is quite promising (Result 1). This result, which demonstrates sustained handwashing over a short period of time, may be evidence that HOPE SOAP<sup>®</sup> can generate behavior change that sticks. Despite this encouraging finding, as noted, the pilot evaluation of HOPE SOAP<sup>®</sup> took place within a relatively short time frame (12 weeks), and therefore observed handwashing at snack-test 2 may not be an accurate representation of caregiver behavior in the long-run. Accordingly, future research should include long-term follow ups with treated caregivers to measure if the intervention's impacts on handwashing behavior are, in fact, sustained in the long-run.

Lastly, although analysis shows that a caregiver's level of interest does not influence how HOPE SOAP<sup>®</sup> affects behavior (Result 5), the fact that the treatment improves caregiver interest in the Family-in-Focus program as a whole is an important result within itself (Result 4). In interpreting this result, it is important to remember how the intervention was designed. Caregivers in the treatment and control groups received exactly the same programming throughout the intervention period, and their children all received the same number of bars of soap and toys. The only difference in the two experimental groups was that control children received plain soap, and treatment children received HOPE SOAP<sup>®</sup>. Therefore, the significant impact that the treatment has on caregiver interest may be directly due to the design of HOPE SOAP<sup>®</sup> itself. Behavioral science notes that positive connotations can lead to behavior change because of humans' sub-conscious desires to associate with positive stimulation (Neal et al. 2015; Marteau, Hollands, and Fletcher 2012). Thus, it may be the case that caregivers place a positive association on HOPE SOAP<sup>®</sup> simply due to its fun and unusual design; and this is enough to encourage more active program participation. This result is important from a policy perspective, as it shows how a small change to an intervention's design can lead caregivers to more actively engage in early childhood development programming. However, there is not enough data available on caregivers' perceptions of HOPE SOAP<sup>®</sup> or the Foundation for Community Work program as a whole to make any definite conclusions about the underlying reasons for their increased responsiveness. In future research, investigators should also collect more detailed qualitative data which can be used to better understand the mechanisms that lead to this behavior change.

In sum, the analysis of the effects of the treatment on caregiver behavior is certainly encouraging. HOPE SOAP<sup>®</sup> shows the promise of improving caregiver handwashing behavior by disrupting existing bad-hygiene habits and strengthening household social norms. The ability of the intervention to disrupt existing behaviors and update social norms (an underlying driver of habitual behavior), coupled with evidence that behavior change due to the treatment is sustained over time, implies that HOPE SOAP<sup>®</sup> may be successful in inducing long-term habitual handwashing behavior.

### 5.3 Limited ability to observe health impacts

Although HOPE SOAP<sup>®</sup> shows great potential to improve caregiver handwashing behavior, the intervention's impacts on household health are negligible. As extrapolated on above, because HOPE SOAP<sup>®</sup> improves the health of treated children, in theory it should also lead to positive health externalities for other household members by reducing overall rates of disease transmission (Burns, Maughan-Brown, and Mouzinho 2017; Chung et al. 2017). However, the analysis conducted above illustrates that a child's assignment to HOPE SOAP<sup>®</sup> has no significant effects on the health outcomes of non-treated individuals within the treated households (Result 6). The null impacts of HOPE SOAP<sup>®</sup> on household health may be explained by a variety of factors. Firstly, results from goodness-of-fit tests signal that the models' specifications are unable to properly predict health outcomes. The low R-squareds in the OLS regressions, in addition to the significant p-values which follow from Hosmer-Lemeshow tests on the binary models (see Appendix 7), show that very little of the variation in health outcomes can be attributed to the models' specifications. Thus, the models do not explain the data well, and can say little about the impact of HOPE SOAP<sup>®</sup> on household member health.

Empirically, there are a few reasons why the models may find no effect of the treatment. One explanation could be due to the fact that the sample population was generally healthy. At baseline and endline, caregivers reported that very few individuals in their households actually had any symptoms of illness. The small absolute number of adverse health outcomes also means that the data lacks variation; and as such, it is harder to tease out if any treatment effects did occur. Moreover, there might actually be no differences in health outcomes between individuals in the treatment and control households. Both of the experimental groups received soap, and so no differences in health outcomes could mean one of three things: (1) both groups experienced no change in health as a result of the intervention; (2) both groups experienced better health due to the provision of soap (regardless if that soap was normal or HOPE SOAP<sup>®</sup>); or (3) both groups experienced worse health than if they would have received no intervention. In the absence of a pure control group (i.e. one that does not receive any soap) there is no way to know which of these three outcomes did occur.

Simply looking at the difference in illness rates for all individuals pre- and post-experiment does show that soap provision may have had some positive health implications for individuals in both groups. From the baseline survey in September to the endline survey in December, the percentage of household members who experienced any symptom of illness declined by roughly 4 percentage points.<sup>43</sup> This decline could be the result of a variety of non-experimentally controlled factors; and, in the absence of a control group, cannot be attributed exclusively to the treatment. However, these improvements in health occurred in spite of the fact that December (the time of endline data collection) is the height of the sickness season in the Western Province of South Africa; when the warm and dry summer climate creates favorable conditions for the spread of illness (Musengimana et al. 2016). Although due to pilot constraints it was not possible to include a pure control group in the evaluation of HOPE SOAP<sup>®</sup>, future research on should attempt to do so in order account for seasonality and enable the accurate assessment of health outcomes over time against a counterfactual.

---

<sup>43</sup> At baseline (endline), 17.72% (13.46%) of individuals report any symptom of illness.

While the possibility that the provision of soap had no impact on the health of individuals in either experimental group is slightly disheartening, this finding would not be inconsistent with the mixed evidence which is present across the existing literature. Many impact evaluations observe no improvements in the health outcomes of individuals who received handwashing interventions (Galiani et al. 2016; Briceño, Coville, and Martinez 2015; Chase and Do 2012). Further, even when studies found improvements to health outcomes, these improvements were typically only observed in the long-term, while in the short-term no impacts were found (Luby et al. 2005). Hence, because health outcomes were only observed at 12-weeks in this experiment, it is also conceivable that the impacts of HOPE SOAP<sup>®</sup> are not non-existent, but simply did not have time to materialize before endline data collection. Future studies should consider conducting long-term follow up surveys in order to measure the impact of handwashing interventions over time.

Another factor which could explain the non-discernable differences in the health outcomes of individual household members in the treatment and control groups is the finding that, in 44% of households where children received HOPE SOAP<sup>®</sup>, children broke open or destroyed the soap to obtain the toy inside.<sup>44</sup> As such, in households where a treated-child “cheated” to retrieve their toy, control group children may actually have had more access to the intervention soap. Evidence from a regression analysis does show that when children cheat to obtain the toy, individuals in HOPE SOAP<sup>®</sup> households experience significantly more flu symptoms than those in control households ( $\beta_T = 0.07$ ,  $pval = 0.09$ ). Conversely, when treated children do not cheat, household members experience significantly less nausea symptoms than those in the control ( $\beta_T = -0.02$ ,  $pval = 0.07$ ).<sup>45,46</sup> These findings provide valuable insights for practitioners and policy makers who are interested in developing successful handwashing interventions. Firstly, the fact that children did not always use HOPE SOAP<sup>®</sup> as intended illustrates that interventions can often lead to unexpected outcomes when implemented in the field. This lesson can be applied to work in any development context. In the case of this intervention specifically, the finding that many children did not use the soap as envisioned may mean that the provision of normal soap could actually be more effective in improving health than HOPE SOAP<sup>®</sup>. Nevertheless, if HOPE SOAP<sup>®</sup> is successful in generating habitual handwashing behavior in households in the treatment group, it still could prove to be a better way to improve health in the long-term, regardless of whether or not a child cheats to uncover the toy at its center. Clearly, more research is needed into unpacking this relationship before any conclusions can be made on the effectiveness of HOPE SOAP<sup>®</sup> in improving health outcomes.

---

<sup>44</sup> This finding is based on a blunt measure because caregivers would have had to both witness and report their child’s cheating behavior; and caregivers were only asked if their child had ever cheated, not about the intensity or instance of cheating.

<sup>45</sup> However, these results should be interpreted with caution. Firstly, low R-squareds present call into question the validity of the models. Second, just as in Table 15, coefficients on the treatment indicators across all health outcomes are small and both positive and negative: meaning that there is not a clear consensus as to the direction of the effect of cheating on health. Finally, Wald tests for joint significance of the indicators return low F-stats and high p-values: despite statistical significance in the models which regress the treatment on nausea and flu symptoms, it still possible that the reported health outcomes of individuals in toy-cheat and non-toy-cheat treatment households are the same.

<sup>46</sup> See Appendix 9 for the relevant regression output.

Lastly, analysis may not be able to discern if there are any differences between the health of individuals in treatment and control groups because the whole sample population resides within the same community. Just as health externalities may cause the intervention's effect to spillover from treated children to non-treated household members, externalities can also exist at the community level. Many published studies document the existence of positive community-wide externalities in relation to health interventions (Andres et al. 2014; Miguel and Kremer 2004; Hawley et al. 2003). Presumably, the positive health effects of HOPE SOAP<sup>®</sup>, which theoretically may exist, could simply spillover from the treatment households to the control households: improving the health of individuals in both groups. However, positive community wide spillovers which result from HOPE SOAP<sup>®</sup> are highly unlikely. Recall that there is not a correlation between caregiver handwashing behavior and better health outcomes in the context of this evaluation. It is therefore evident that handwashing behavior is not a major determinant of health outcomes in the evaluation context. Alternatively, other more influential determinants of health, such as community-wide sanitation and hygiene problems (open defecation, poor water quality, etc.), may lead to negative externalities for households in the entire sample. Negative community-wide externalities could dilute the effects of the intervention and/or weaken the ability of the analysis to detect any significant treatment effects (Andres et al. 2014; Briceño, Coville, and Martinez 2015). Again, this possibility should serve as a lesson for future research into the relationship between handwashing and health: investigators should design experiments that control for, and include indicators to measure, both the effects of spillovers due to health externalities, and the overall determinants of health outcomes within an evaluation's context.

## 6. CONCLUSION

In conclusion, this paper makes a few significant contributions to the existing literature on handwashing behavior change interventions. Firstly, few studies which examine the effects of handwashing promotion programs on behavior and health actually take place in the field, in low and middle-income country contexts; and those that do provide little concrete evidence about the ability of handwashing encouragement programs to improve handwashing or health in practice. This may be because evaluated-interventions rarely consider the habitual nature of handwashing behaviors. In contrast, this paper presents an evaluation of HOPE SOAP<sup>®</sup>, an intervention which specifically targets the drivers of habitual behaviors, and is delivered to urban-poor households in a developing country. Second, while spillover effects undeniably occur during the provision health interventions, no studies to date examine if health externalities, behavioral spillovers, or interaction effects are present in the context of handwashing promotion programs. Conversely, this paper solely focuses on the effects that a child receiving HOPE SOAP<sup>®</sup> has on the handwashing behaviors and health of non-targeted individuals in their households.

The analysis offered in this paper illustrates that a child's assignment to HOPE SOAP<sup>®</sup> treatment leads to a positive impact on the handwashing behaviors of their caregiver. By disrupting existing poor-hygiene habits and strengthening handwashing social norms within households, HOPE SOAP<sup>®</sup> has the potential to create handwashing behavior change which may be sustained in the long-run. Furthermore, evidence indicates that the effect of HOPE SOAP<sup>®</sup> on caregiver handwashing behavior may occur through spillovers from treated children. This finding reveals the potential for behavioral spillovers to transpire as a result of handwashing interventions: serving as a consideration for policy makers and practitioners when developing effective handwashing interventions, and as a call for more research into spillovers in handwashing more generally. Despite its undeniable positive influence on caregiver handwashing behavior, HOPE SOAP<sup>®</sup> does not lead to positive impacts on the health of individual members of treated households. While this finding seems discouraging at first, there are many statistical and experimental reasons why no effects are detected. As such, this result should rather be thought of as a lesson for future researchers when considering how to better design experiments to measure the impacts of handwashing on health. Nevertheless, although HOPE SOAP<sup>®</sup>'s positive effect on handwashing behavior does not lead to any health impacts in the short-term, because it is successful at inducing a habit of handwashing behavior, in the long-run it may prove a valuable tool to decrease diarrheal disease and ARI: two major causes of child mortality in the developing world.

## REFERENCES

- Andres, Luis A., Bertha Briceno, Claire Chase, and Juan A. Echenique. 2014. "Sanitation and Externalities: Evidence from Early Childhood Health in Rural India." <http://agris.fao.org/agris-search/search.do?recordID=US2014600407>.
- Angelucci, Manuela, and Vincenzo Di Maro. 2010. "Program Evaluation and Spillover Effects." Inter-American Development Bank. <https://publications.iadb.org/bitstream/handle/11319/3457/Program%20Evaluation%20and%20Spillover%20Effects.pdf?sequence=1>.
- Angrist, Joshua, and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton, New Jersey: Princeton University Press.
- Aunger, Robert, Wolf-Peter Schmidt, Ashish Ranpura, Yolande Coombes, Peninnah Mukiri Maina, Carol Nkatha Matiko, and Valerie Curtis. 2010. "Three Kinds of Psychological Determinants for Hand-Washing Behaviour in Kenya." *Social Science & Medicine* 70 (3):383–91. <https://doi.org/10.1016/j.socscimed.2009.10.038>.
- Baron, Reuben M., and David A. Kenny. 1986. "The Moderator–mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51 (6):1173.
- Biran, Adam, Wolf-Peter Schmidt, Richard Wright, Therese Jones, M Seshadri, Pradeep Isaac, N. A. Nathan, et al. 2009. "The Effect of a Soap Promotion and Hygiene Education Campaign on Handwashing Behaviour in Rural India: A Cluster Randomised Trial: The Effect of a Soap Promotion and Hygiene Education Campaign on Handwashing Behaviour in Rural India." *Tropical Medicine & International Health* 14 (10):1303–14. <https://doi.org/10.1111/j.1365-3156.2009.02373.x>.
- Black, Robert E., Simon Cousens, Hope L. Johnson, Joy E. Lawn, Igor Rudan, Diego G. Bassani, Prabhat Jha, et al. 2010. "Global, Regional, and National Causes of Child Mortality in 2008: A Systematic Analysis." *Lancet (London, England)* 375 (9730):1969–87. [https://doi.org/10.1016/S0140-6736\(10\)60549-1](https://doi.org/10.1016/S0140-6736(10)60549-1).
- Black, Robert E., Saul S. Morris, and Jennifer Bryce. 2003. "Where and Why Are 10 Million Children Dying Every Year?" *Lancet (London, England)* 361 (9376):2226–34. [https://doi.org/10.1016/S0140-6736\(03\)13779-8](https://doi.org/10.1016/S0140-6736(03)13779-8).
- Bleakley, Hoyt. 2010. "Health, Human Capital, and Development." *Annual Review of Economics* 2 (September):283–310. <https://doi.org/10.1146/annurev.economics.102308.124436>.
- Bloom, David E., and David Canning. 2008. "Population Health and Economic Growth." Working Paper 24. Washington D.C.: The International Bank for Reconstruction and Development / The World Bank.
- Bloom, David E., David Canning, and Dean T. Jamison. 2004. "Health, Wealth, and Welfare." In *Health and Development*. Washington D.C.: International Monetary Fund.
- Bloom, David E., David Canning, and Jaypee Sevilla. 2004. "The Effect of Health on Economic Growth: A Production Function Approach." *World Development* 32 (1):1–13. <https://doi.org/10.1016/j.worlddev.2003.07.002>.
- Borghi, J., L. Guinness, J. Ouedraogo, and V. Curtis. 2002. "Is Hygiene Promotion Cost-Effective? A Case Study in Burkina Faso." *Tropical Medicine & International Health* 7 (11):960–69. <https://doi.org/10.1046/j.1365-3156.2002.00954.x>.
- Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1):63–82. <https://doi.org/10.1093/pan/mpi014>.
- Briceno, Bertha, Aidan Coville, and Sebastian Martinez. 2015. "Promoting Handwashing and Sanitation: Evidence from a Large-Scale Randomized Trial in Rural Tanzania." Policy Research Working Paper 7164. Washington D.C.: The World Bank.

- Brookes, Sara T., Elise Whitely, Matthias Egger, George Davey Smith, Paul A. Mulheran, and Tim J. Peters. 2004. "Subgroup Analyses in Randomized Trials: Risks of Subgroup-Specific Analyses; Power and Sample Size for the Interaction Test." *Journal of Clinical Epidemiology* 57 (3):229–36. <https://doi.org/10.1016/j.jclinepi.2003.08.009>.
- Burns, Justine, Brendan Maughan-Brown, and Áurea Mouzinho. 2017. "Washing with Hope: Evidence from a Hand-Washing Pilot Study among Children in South Africa." A Southern African Labour and Development Unit Working Paper Number 199. SALDRU, University of Cape Town. [http://opensaldru.uct.ac.za/bitstream/handle/11090/858/2017\\_199\\_Saldruwp.pdf?sequence=1](http://opensaldru.uct.ac.za/bitstream/handle/11090/858/2017_199_Saldruwp.pdf?sequence=1)
- Cairncross, Sandy, Caroline Hunt, Sophie Boisson, Kristof Bostoen, Valerie Curtis, Isaac CH Fung, and Wolf-Peter Schmidt. 2010. "Water, Sanitation and Hygiene for the Prevention of Diarrhoea." *International Journal of Epidemiology* 39 (suppl\_1):i193–205. <https://doi.org/10.1093/ije/dyq035>.
- Cairncross, Sandy, and Vivian Valdmanis. 2006. "Water Supply, Sanitation, and Hygiene Promotion." In *Diseases Control Priorities in Developing Countries*, 2nd ed. Washington D.C.: World Bank.
- Cameron, Lisa A., Manisha Shah, and Susan Olivia. 2013. "Impact Evaluation of a Large-Scale Rural Sanitation Project in Indonesia." Policy Research Working Paper 6360. Impact Evaluation Series No. 83. The World Bank. <https://doi.org/10.1596/1813-9450-6360>.
- Caroli, Eve, and Lexane Weber-Baghdiguian. 2016. "Self-Reported Health and Gender: The Role of Social Norms." *Social Science & Medicine* 153 (March):220–29. <https://doi.org/10.1016/j.socscimed.2016.02.023>
- Chase, Claire, and Quy-Toan Do. 2010. "Scaling Up Handwashing Behavior: Findings from the Impact Evaluation Baseline Survey in Vietnam." Water and Sanitation Program at the World Bank. [https://www.wsp.org/sites/wsp.org/files/publications/WSP\\_VietnamBaselineReport\\_HWWS.pdf](https://www.wsp.org/sites/wsp.org/files/publications/WSP_VietnamBaselineReport_HWWS.pdf).
- Chase, Claire and Quy-Toan Do. 2012. "Handwashing Behavior Change at Scale: Evidence from a Randomized Evaluation in Vietnam." World Bank Policy Research Working Paper 6207. Washington D.C.: The World Bank.
- Christensen, Garret S., and Edward Miguel. 2016. "Transparency, Reproducibility, and the Credibility of Economics Research." Working Paper 22989. National Bureau of Economic Research. <http://www.nber.org/papers/w22989>.
- Chung, Jade, Jaynal Abedin, David Berger, Ashley Clark, Lauren Falcao, Veronica Jimenez, Eugene Konagaya, Diana Tran, Benjamin F. Arnold, and Alan Hubbard. 2015. "The Identification and Measurement of Health-Related Spillovers in Impact Evaluations: A Systematic Review." *Grantee Final Review. New Delhi: 3ie*.
- Chung, Jade, Jaynal Abedin, David Berger, Ashley Clark, Veronica Jimenez, Eugene Konagaya, Diana Tran, et al. 2017. "Spillover Effects on Health Outcomes in Low- and Middle-Income Countries: A Systematic Review." *International Journal of Epidemiology* 46 (4):1251–76. <https://doi.org/10.1093/ije/dyx039>.
- Clift, Jeremy, ed. 2004. *Health and Development: A Compilation of Articles from Finance & Development*. Washington, D.C: International Monetary Fund.
- Collins, Francis. 2015. "Growing Importance of Health in the Economy." *World Economic Forum: Top 10 Trends of 2015* (blog). 2015. <https://widgets.weforum.org/outlook15/10.html>.
- Curtis, Valerie, and Sandy Cairncross. 2003. "Effect of Washing Hands with Soap on Diarrhoea Risk in the Community: A Systematic Review." *The Lancet. Infectious Diseases* 3 (5):275–81.
- Curtis, Valerie A., Lisa O. Danquah, and Robert V. Aunger. 2009. "Planned, Motivated and Habitual Hygiene Behaviour: An Eleven Country Review." *Health Education Research* 24 (4):655–73. <https://doi.org/10.1093/her/cyp002>.
- Deaton, Angus. 1997. "The Analysis of Household Surveys : A Microeconomic Approach to Development Policy." 17140. The World Bank.

- <http://documents.worldbank.org/curated/en/593871468777303124/The-analysis-of-household-surveys-a-microeconomic-approach-to-development-policy>.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2006. "Using Randomization in Development Economics Research: A Toolkit." Working Paper 333. National Bureau of Economic Research. <http://www.nber.org/papers/t0333>.
- Dupas, Pascaline. 2011. "Health Behavior in Developing Countries." *Annual Review of Economics* 3 (September):425–99.
- Ejemot, Regina I., John E. Ehiri, Martin M. Meremikwu, and Julia A. Critchley. 2008. "Hand Washing for Preventing Diarrhoea." *The Cochrane Database of Systematic Reviews*, no. 1 (January):CD004265. <https://doi.org/10.1002/14651858.CD004265.pub2>.
- Galiani, Sebastian, Paul Gertler, Nicolas Ajzenman, and Alexandra Orsola-Vidal. 2016. "Promoting Handwashing Behavior: The Effects of Large-Scale Community and School-Level Interventions." *Health Economics* 25 (12):1545–59. <https://doi.org/10.1002/hec.3273>.
- Galiani, Sebastian, Paul Gertler, Alexandra Orsola-Vidal. 2012. *Promoting Handwashing Behavior in Peru: The Effect of Large-Scale Mass-Media and Community Level Interventions*. Policy Research Working Papers. The World Bank. <http://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-6257>.
- Gelman, Andrew, Jennifer Hill, and Masanao Yajima. 2012. "Why We (Usually) Don't Have to Worry About Multiple Comparisons." *Journal of Research on Educational Effectiveness* 5 (2):189–211. <https://doi.org/10.1080/19345747.2011.618213>.
- Gertler, Paul, Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M.J. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington D.C.: The World Bank.
- Glennerster, Rachel, and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton, New Jersey: Princeton University Press.
- Glewwe, Paul, and Edward A. Miguel. 2007. "The Impact of Child Health and Nutrition on Education in Less Developed Countries." In *Handbook of Development Economics*, 4:3561–3606. Elsevier. <http://linkinghub.elsevier.com/retrieve/pii/S1573447107040569>.
- Glewwe, Paul, and Karthik Muralidharan. 2015. "Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." *University of Oxford, Research on Improving Systems of Education (RISE)*.
- Greenland, Sander, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman. 2016. "Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations." *European Journal of Epidemiology* 31:337–50. <https://doi.org/10.1007/s10654-016-0149-3>.
- Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *The Quarterly Journal of Economics*, July, qjw025. doi:10.1093/qje/qjw025.
- Hawley, William A., Penelope A. Phillips-Howard, Feiko O. Ter Kuile, Dianne J. Terlouw, John M. Vulule, Maurice Ombok, Bernard L. Nahlen, et al. 2003. "Community-Wide Effects of Permethrin-Treated Bed Nets on Child Mortality and Malaria Morbidity in Western Kenya." *The American Journal of Tropical Medicine and Hygiene* 68 (4\_suppl):121–27. <https://doi.org/10.4269/ajtmh.2003.68.121>.
- Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics*, 65–70.
- Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol. 2016. "Habit Formation and Rational Addiction: A Field Experiment in Handwashing." Working Paper. Yale University, Dhaka University, MIT, Harvard University.
- Hyder, Adnan A., Prasanthi Puvanachandra, and Richard H. Morrow. 2012. "Measuring the Health of Populations: Explaining Composite Indicators." *Journal of Public Health Research* 1 (3): 222–28. <https://doi.org/10.4081/jphr.2012.e35>.

- Kazdin, Alan E. 2007. "Mediators and Mechanisms of Change in Psychotherapy Research." *Annual Review of Clinical Psychology* 3 (1):1–27. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091432>.
- Lawson, Peter J., and Susan A. Flocke. 2009. "Teachable Moments for Health Behavior Change: A Concept Analysis." *Patient Education and Counseling* 76 (1):25–30. <https://doi.org/10.1016/j.pec.2008.11.002>.
- Lee, Soohyung, and Azeem M. Shaikh. 2014. "Multiple Testing and Heterogenous Treatment Effects: Re-Evaluating the Effect of PROGRESA on School Enrollment: Heterogenous Treatment Effects." *Journal of Applied Econometrics* 29 (4):612–26. <https://doi.org/10.1002/jae.2327>.
- Lopez, Alan D., Colin D. Mathers, Majid Ezzati, Dean T. Jamison, and Christopher J. L. Murray. 2006. "Global and Regional Burden of Disease and Risk Factors, 2001: Systematic Analysis of Population Health Data." *Lancet (London, England)* 367 (9524):1747–57. [https://doi.org/10.1016/S0140-6736\(06\)68770-9](https://doi.org/10.1016/S0140-6736(06)68770-9).
- Luby, Stephen P., Mubina Agboatwalla, Daniel R. Feikin, John Painter, Ward Billhimer, Arshad Altaf, and Robert M. Hoekstra. 2005. "Effect of Handwashing on Child Health: A Randomised Controlled Trial." *Lancet (London, England)* 366 (9481):225–33. [https://doi.org/10.1016/S0140-6736\(05\)66912-7](https://doi.org/10.1016/S0140-6736(05)66912-7).
- Marteau, Theresa M., Gareth J. Hollands, and Paul C. Fletcher. 2012. "Changing Human Behavior to Prevent Disease: The Importance of Targeting Automatic Processes." *Science (New York, N.Y.)* 337 (6101):1492–95. <https://doi.org/10.1126/science.1226918>.
- Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1):159–217. <https://doi.org/10.1111/j.1468-0262.2004.00481.x>.
- Musengimana, Gentile, Fidele K. Mukinda, Roderick Machekano, and Hassan Mahomed. 2016. "Temperature Variability and Occurrence of Diarrhoea in Children under Five-Years-Old in Cape Town Metropolitan Sub-Districts." *International Journal of Environmental Research and Public Health* 13 (9). <https://doi.org/10.3390/ijerph13090859>.
- Neal, David T., Wendy Wood, Jennifer S. Labrecque, and Phillippa Lally. 2012. "How Do Habits Guide Behavior? Perceived and Actual Triggers of Habits in Daily Life." *Journal of Experimental Social Psychology* 48 (2):492–98. <https://doi.org/10.1016/j.jesp.2011.10.011>.
- Neal, David, Jelena Vujcic, Orlando Hernandez, and Wendy Wood. 2015. "The Science of Habit: Creating Disruptive and Sticky Behavior Change in Handwashing Behavior." Washington D.C.: USAID/WASHplus Project. [http://www.washplus.org/sites/default/files/resource\\_files/habits-neal2015.pdf](http://www.washplus.org/sites/default/files/resource_files/habits-neal2015.pdf).
- Orsola-Vidal, Alexandra, and Ahmad Yusuf. 2011. "Scaling Up Handwashing Behavior: Findings from the Impact Evaluation Baseline Survey in Senegal." Water and Sanitation Program at the World Bank.
- Ouellette, Judith A., and Wendy Wood. 1998. "Habit and Intention in Everyday Life: The Multiple Processes by Which Past Behavior Predicts Future Behavior." *Psychological Bulletin* 124 (1):54–74.
- Prüss-Ustün, Annette, David Kay, Lorna Fewtrell, and Jamie Bartram. 2002. "Estimating the Burden of Disease from Water, Sanitation, and Hygiene at a Global Level." *Environmental Health Perspectives* 110 (5):537.
- Prüss-Ustün, Annette, Jamie Bartram, Thomas Clasen, John M. Colford, Oliver Cumming, Valerie Curtis, Sophie Bonjour, et al. 2014. "Burden of Disease from Inadequate Water, Sanitation and Hygiene in Low- and Middle-Income Settings: A Retrospective Analysis of Data from 145 Countries." *Tropical Medicine & International Health: TM & IH* 19 (8):894–905. <https://doi.org/10.1111/tmi.12329>.
- Rabie, Tamer, and Valerie Curtis. 2006. "Handwashing and Risk of Respiratory Infections: A Quantitative Systematic Review." *Tropical Medicine & International Health: TM & IH* 11 (3):258–67. <https://doi.org/10.1111/j.1365-3156.2006.01568.x>.

- Ram, Pavani K. 2010. "Practical Guidance for Measuring Handwashing Behavior." Water and Sanitation Program: Working Paper. Washington D.C.: Water and Sanitation Program at the World Bank.
- Rosen, Sydney, and Jeffrey R. Vincent. 1999. *Household Water Resources and Rural Productivity in Sub-Saharan Africa: A Review of the Evidence*. Vol. 673. Harvard Institute for International Development, Harvard University Cambridge, MA.  
<http://www.cid.harvard.edu/archive/events/cidneudc/papers/rosenvincent.pdf>.
- Schmidt, Wolf-Peter. 2014. "The Elusive Effect of Water and Sanitation on the Global Burden of Disease." *Tropical Medicine & International Health* 19 (5):522–27. <https://doi.org/10.1111/tmi.12286>.
- Statistics South Africa. 2017a. "Living Conditions of Households in South Africa: An Analysis of Household Expenditure and Income Data Using the LCS 2014/2015." Statistical Release P0310. Statistics South Africa. <http://www.statssa.gov.za/publications/P0310/P03102014.pdf>.
- Statistics South Africa. 2017b. "General Household Survey: 2016." Statistical Release P0318. Pretoria: Statistics South Africa. <http://www.statssa.gov.za/publications/P0318/P03182016.pdf>.
- Taylor, Dawn L., Tanya M. Kahawita, Sandy Cairncross, and Jeroen H. J. Ensink. 2015. "The Impact of Water, Sanitation and Hygiene Interventions to Control Cholera: A Systematic Review." *PLOS ONE* 10 (8):e0135676. <https://doi.org/10.1371/journal.pone.0135676>.
- United Nations Children's Fund. 2014. "Levels & Trends in Child Mortality." Report 2014. New York, NY: United Nations Children's Fund.  
[https://www.unicef.org/media/files/Levels\\_and\\_Trends\\_in\\_Child\\_Mortality\\_2014.pdf](https://www.unicef.org/media/files/Levels_and_Trends_in_Child_Mortality_2014.pdf).
- Vindigni, Stephen M., Patricia L. Riley, and Michael Jhung. 2011. "Systematic Review: Handwashing Behaviour in Low- to Middle-Income Countries: Outcome Measures and Behaviour Maintenance." *Tropical Medicine & International Health: TM & IH* 16 (4):466–77. <https://doi.org/10.1111/j.1365-3156.2010.02720.x>.
- Wooldridge, Jeffrey M. 2009. *Introductory Econometrics*. Fourth. South-Western, a part of Cengage Learning.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Second Edition. The MIT Press.
- World Bank. 2005. "The Handwashing Handbook: A Guide for Developing a Hygiene Promotion Program to Increase Handwashing with Soap." 32302. The World Bank.  
<http://documents.worldbank.org/curated/en/681501468141299225/The-handwashing-handbook-a-guide-for-developing-a-hygiene-promotion-program-to-increase-handwashing-with-soap>.
- World Health Organization, ed. 2009. *Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks*. Geneva, Switzerland: World Health Organization.
- World Health Organization. 2017. "Global Health Observatory Data Repository."  
<http://apps.who.int/gho/data/view.main.ghe3002015-CH3?lang=en>.
- Zwane, Alix Peterson, and Michael Kremer. 2007. "What Works in Fighting Diarrheal Diseases in Developing Countries? A Critical Review." Working Paper No. 140. Harvard University: Center for International Development.  
[http://www.environment.harvard.edu/docs/faculty\\_pubs/kremer\\_fighting.pdf](http://www.environment.harvard.edu/docs/faculty_pubs/kremer_fighting.pdf).

## APPENDIX

Appendix 1: Child behavior as a mediator for HOPE SOAP<sup>®</sup>'s effect on caregiver behavior  
(OLS/Logit/Probit results)

VARIABLES	(1) OLS	(2) Logit	(3) Logit APE	(4) Probit	(5) Probit APE
<b>Snack-test 1</b>					
Treatment coefficient	-0.15***	-1.64**	-0.13**	-0.86**	-0.13**
<i>Standard error</i>	(0.05)	(0.65)	(0.05)	(0.33)	(0.05)
<i>P-value</i>	0.01	0.01	0.01	0.01	0.01
Child wash at snack-test 1	0.75***	4.73***	0.39***	2.67***	0.42***
	(0.06)	(0.73)	(0.03)	(0.35)	(0.02)
	0.00	0.00	0.00	0.00	0.00
R <sup>2</sup> / Hosmer-Lemeshow pval	0.61	0.00		0.00	
Observations	143	146	146	146	146
Controls	Yes	Yes	Yes	Yes	Yes
<b>Snack-test 2</b>					
Treatment	0.05	0.77	0.06	0.38	0.05
	(0.05)	(0.69)	(0.05)	(0.34)	(0.05)
	0.32	0.27	0.26	0.28	0.27
Child wash at snack-test 2	0.75***	4.61***	0.38***	2.63***	0.41***
	(0.05)	(0.67)	(0.02)	(0.32)	(0.02)
	0.00	0.00	0.00	0.00	0.00
R <sup>2</sup> / Hosmer-Lemeshow pval	0.65	0.00		0.00	
Observations	118	118	118	118	118
Controls	Yes	Yes	Yes	Yes	Yes
<b>Snack-test 2</b>					
Treatment	0.05	0.76	0.06	0.36	0.05
	(0.05)	(0.68)	(0.05)	(0.33)	(0.05)
	0.32	0.27	0.26	0.28	0.27
Child wash at snack-test 2	0.77***	4.94***	0.41***	2.73***	0.42***
	(0.06)	(1.09)	(0.06)	(0.47)	(0.04)
	0.00	0.00	0.00	0.00	0.00
Snack-test 1	-0.02	-0.57	-0.04	-0.20	-0.03
	(0.05)	(0.95)	(0.07)	(0.45)	(0.06)
	0.65	0.55	0.54	0.67	0.66
R <sup>2</sup> / Hosmer-Lemeshow pval	0.65	0.00		0.00	
Observations	118	118	118	118	118
Controls	Yes	Yes	Yes	Yes	Yes

<b>Both snack-tests</b>					
Treatment	0.02 (0.07) 0.77	0.23 (0.68) 0.74	0.02 (0.06) 0.73	0.08 (0.34) 0.82	0.01 (0.06) 0.82
Child wash at both snack-tests	0.65*** (0.07) 0.00	3.64*** (0.59) 0.00	0.37*** (0.01) 0.00	2.11*** (0.30) 0.00	0.39*** (0.02) 0.00
R <sup>2</sup> / Hosmer-Lemeshow pval	0.55	0.00		0.00	
Observations	116	118	118	118	118
Controls	Yes	Yes	Yes	Yes	Yes

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix 2: Effect of HOPE SOAP<sup>®</sup> on caregiver behavior for a subpopulation of caregivers from households who were unclean at baseline (OLS results)**

VARIABLES	(1)	(2)	(3)
	Snack-test 1	Snack-test 2	Both
Treatment	-0.01 (0.12) 0.90	0.16 (0.11) 0.14	0.14 (0.10) 0.16
Baseline cleanliness	0.16 (0.13) 0.23	-0.10 (0.13) 0.44	-0.05 (0.12) 0.65
Piped water in house	0.08*** (0.02) 0.00	0.08*** (0.02) 0.00	0.10*** (0.02) 0.00
Household limits water use	0.02 (0.10) 0.87	0.02 (0.10) 0.87	0.07 (0.10) 0.49
Constant	0.10 (0.17) 0.57	0.02 (0.16) 0.92	-0.18 (0.14) 0.19
Observations	169	138	138
R-squared	0.15	0.10	0.18
Controls	Yes	Yes	Yes

Standard errors in parentheses, followed by pvals

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix 3: Differential effects of HOPE SOAP<sup>®</sup> by baseline household cleanliness  
(OLS/Logit/Probit results)**

VARIABLES	(1) OLS	(2) Logit	(3) Logit APE	(4) Probit	(5) Probit APE
<b>Snack-test 1</b>					
Treatment coefficient	0.01	0.04	-0.00	0.03	-0.00
<i>Standard error</i>	(0.12)	(0.51)	(0.11)	(0.32)	(0.11)
<i>P-value</i>	0.92	0.93	0.93	0.93	0.93
Clean household	0.35*** (0.10) 0.00	1.64*** (0.52) 0.00	0.35*** (0.10) 0.00	1.00*** (0.31) 0.00	0.35*** (0.09) 0.00
Clean*Treatment	-0.23 (0.15) 0.12	-1.15 (0.70) 0.10	-0.24* (0.14) 0.09	-0.69 (0.42) 0.11	-0.24* (0.14) 0.09
R <sup>2</sup> / Hosmer-Lemeshow pval	0.09	0.90		0.90	
Observations	171	171	171	171	171
Controls	Yes	Yes	Yes	Yes	Yes
<b>Snack-test 2</b>					
Treatment	0.19 (0.12) 0.12	0.96 (0.63) 0.13	0.21 (0.13) 0.11	0.55 (0.37) 0.14	0.19 (0.12) 0.12
Clean household	0.35*** (0.12) 0.00	1.68** (0.66) 0.01	0.36*** (0.12) 0.00	0.99** (0.38) 0.01	0.35*** 0.12 0.00
Clean*Treatment	-0.21 (0.17) 0.20	-1.10 (0.82) 0.18	-0.24 (0.17) 0.16	-0.63 (0.48) 0.20	-0.22 (0.16) 0.18
R <sup>2</sup> / Hosmer-Lemeshow pval	0.10	0.83		0.81	
Observations	139	139	139	139	139
Controls	Yes	Yes	Yes	Yes	Yes
<b>Snack-test 2</b>					
Treatment	0.15 (0.10) 0.15	0.88 (0.64) 0.17	0.16 (0.11) 0.15	0.49 (0.38) 0.20	0.15 (0.11) 0.18
Clean Household	0.19 (0.13) 0.14	1.05 (0.69) 0.13	0.19 (0.12) 0.11	0.60 (0.41) 0.14	0.19 (0.12) 0.13
Clean*Treatment	-0.11 (0.16) 0.49	-0.68 (0.86) 0.43	-0.12 (0.15) 0.42	-0.36 (0.51) 0.47	-0.11 (0.15) 0.47

Snack-test 1	0.38*** (0.08) 0.00	1.82*** (0.43) 0.00	0.34*** (0.05) 0.00	1.09*** (0.25) 0.00	0.34*** (0.06) 0.00
R <sup>2</sup> / Hosmer-Lemeshow pval	0.23	0.97		0.95	
Observations	139	119	119	119	119
Controls	Yes	Yes	Yes	Yes	Yes
<b>Both snack-tests</b>					
Treatment	0.16 (0.10) 0.12	1.06 (0.70) 0.13	0.21 (0.12) 0.11	0.58 (0.39) 0.14	0.19 (0.12) 0.13
Clean Household	0.36*** (0.11) 0.00	2.00*** (0.72) 0.01	0.39*** (0.12) 0.00	1.16*** (0.40) 0.00	0.38*** (0.11) 0.00
Clean*Treatment	-0.22 (0.15) 0.15	-1.40 (0.89) 0.12	-0.27 (0.12) 0.10	-0.78 (0.51) 0.13	-0.25 (0.16) 0.11
R <sup>2</sup> / Hosmer-Lemeshow pval	0.13	0.97		0.97	
Observations	139	139	119	139	119
Controls	Yes	Yes	Yes	Yes	Yes

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

#### Appendix 4: Caregiver responsiveness variables

<i>Variable</i>	<i>Survey Question(s)</i>	<i>Instrument type</i>	<i>Response type</i>
Caregiver interest	Please rate...’s interest in today’s lesson on a scale of 1 to 10.	Observed by FCW health worker at FIF visit	Discrete number between 1 and 10
Number of FCW sessions attended	In the last two months, has your FCW home-visitor invited you or made you aware of any parent workshops?	Reported by caregiver at endline survey	Dummy variable, equal to 1 if yes
	[If yes, then] How many of these workshops did you attend?	-	Discrete number between 0 and 10

**Appendix 5: Interest as an additional covariate in the regression of the treatment on behavior  
(OLS/Logit/Probit results)**

VARIABLES	(1) OLS	(2) Logit	(3) Logit APE	(4) Probit	(5) Probit APE
<b>Snack-test 1</b>					
Treatment coefficient	-0.11	-0.63	-0.11	-0.39	-0.11
<i>Standard error</i>	(0.09)	(0.53)	(0.09)	(0.31)	(0.09)
<i>P-value</i>	0.23	0.24	0.22	0.22	0.20
Interest	0.01	0.05	0.00	0.03	0.00
	(0.02)	(0.11)	(0.02)	(0.07)	(0.02)
	0.64	0.66	0.66	0.63	0.63
R <sup>2</sup> / Hosmer-Lemeshow pval	0.24	0.00		0.00	
Observations	99	99	99	99	99
Controls	Yes	Yes	Yes	Yes	Yes
<b>Snack-test 2</b>					
Treatment	0.14	0.61	0.13	0.39	0.14
	(0.11)	(0.46)	(0.10)	(0.29)	(0.10)
	0.19	0.19	0.17	0.18	0.16
Interest	0.00	0.00	0.00	0.00	0.00
	(0.02)	(0.10)	(0.02)	(0.06)	(0.02)
	0.95	0.99	0.98	1.00	0.99
R <sup>2</sup> / Hosmer-Lemeshow pval	0.09	0.19		0.19	
Observations	95	95	95	95	95
Controls	Yes	Yes	Yes	Yes	Yes
<b>Snack-test 2</b>					
Treatment	0.21**	1.21**	0.20**	0.69**	0.20**
	(0.09)	(0.58)	(0.09)	(0.33)	(0.08)
	0.02	0.04	0.02	0.04	0.02
Interest	-0.00	-0.02	-0.00	-0.00	-0.00
	(0.02)	(0.12)	(0.02)	(0.07)	(0.02)
	0.85	0.88	0.87	0.95	0.95
Snack-test 1	0.55***	2.73***	0.46***	1.62***	0.47
	(0.09)	(0.61)	(0.05)	(0.35)	(0.06)
	0.00	0.00	0.00	0.00	0.00
R <sup>2</sup> / Hosmer-Lemeshow pval	0.30	0.00		0.88	
Observations	89	95	95	95	95
Controls	Yes	Yes	Yes	Yes	Yes
<b>Both snack-tests</b>					
Treatment	0.10	0.50	0.10	0.32	0.10
	(0.10)	(0.48)	(0.09)	(0.29)	(0.09)
	0.32	0.30	0.29	0.28	0.27

Interest	-0.01 (0.02) 0.63	-0.08 (0.11) 0.50	-0.01 (0.02) 0.50	-0.04 (0.07) 0.53	-0.01 (0.02) 0.52
R <sup>2</sup> / Hosmer-Lemeshow pval	0.16	0.97		0.94	
Observations	95	95	95	95	95
Controls	Yes	Yes	Yes	Yes	Yes

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix 6: Knowledge as an additional covariate in the regression of the treatment on behavior  
(OLS/Logit/Probit results)**

VARIABLES	(1) OLS	(2) Logit	(3) Logit APE	(4) Probit	(5) Probit APE
<b>Snack-test 1</b>					
Treatment coefficient	-0.09	-0.47	-0.09	-0.29	-0.09
<i>Standard error</i>	(0.08)	(0.42)	(0.08)	(0.25)	(0.08)
<i>P-value</i>	0.27	0.27	0.26	0.24	0.23
Knowledge	0.02	0.10	0.02	0.06	0.02
	(0.02)	(0.09)	(0.01)	(0.06)	(0.01)
	0.28	0.26	0.24	0.27	0.25
R <sup>2</sup> / Hosmer-Lemeshow pval	0.14	0.06		0.24	
Observations	145	145	145	145	145
Controls	Yes	Yes	Yes	Yes	Yes
<b>Snack-test 2</b>					
Treatment	0.13	0.58	0.12	0.37	0.12
	(0.09)	(0.42)	(0.08)	(0.26)	(0.08)
	0.17	0.17	0.16	0.15	0.14
Knowledge	0.00	0.01	0.00	0.01	0.00
	(0.02)	(0.10)	(0.02)	(0.06)	(0.02)
	0.90	0.89	0.89	0.88	0.87
R <sup>2</sup> / Hosmer-Lemeshow pval	0.14	0.08		0.09	
Observations	119	119	119	119	119
Controls	Yes	Yes	Yes	Yes	Yes
<b>Snack-test 2</b>					
Treatment	0.14*	0.79*	0.13*	0.47*	0.14*
	(0.08)	(0.45)	(0.07)	(0.27)	(0.07)
	0.08	0.09	0.07	0.08	0.07
Knowledge	-0.01	-0.07	-0.01	-0.04	-0.01
	(0.02)	(0.11)	(0.01)	(0.06)	(0.01)
	0.48	0.51	0.50	0.51	0.50
Snack-test 1	0.44***	2.20***	0.38***	1.30***	0.39***
	(0.08)	(0.48)	(0.05)	(0.27)	(0.06)
	0.00	0.00	0.00	0.00	0.00
R <sup>2</sup> / Hosmer-Lemeshow pval	0.29	0.00		0.00	
Observations	119	119	119	119	119
Controls	Yes	Yes	Yes	Yes	Yes
<b>Both snack-tests</b>					
Treatment	0.07	0.35	0.06	0.23	0.07
	(0.09)	(0.44)	(0.08)	(0.26)	(0.08)
	0.42	0.42	0.42	0.39	0.38

Knowledge	0.02 (0.02) 0.41	0.08 (0.12) 0.48	0.01 (0.08) 0.47	0.05 (0.07) 0.48	0.01 (0.02) 0.47
R <sup>2</sup> / Hosmer-Lemeshow pval	0.19	0.26		0.60	
Observations	119	119	119	119	119
Controls	Yes	Yes	Yes	Yes	Yes

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix 7: Effect of HOPE SOAP<sup>®</sup> on household member health (OLS/Logit/Probit results)**

VARIABLE	(1) OLS	(2) Logit	(3) Logit APE	(4) Probit	(5) Probit APE
<b>Diarrhea</b>					
Treatment	-0.00	-0.08	-0.00	-0.04	-0.00
<i>Standard error</i>	(0.01)	(0.43)	(0.01)	(0.19)	(0.01)
<i>P-value</i>	0.84	0.84	0.84	0.84	0.84
R <sup>2</sup> / Hosmer-Lemeshow pval	0.00				
Observations	773	773	773	773	773
Controls	Yes	Yes	Yes	Yes	Yes
<b>Flu</b>					
Treatment	0.04	0.50	0.04	0.26	0.04
	(0.03)	(0.36)	(0.03)	(0.18)	(0.03)
	0.18	0.17	0.18	0.16	0.16
R <sup>2</sup> / Hosmer-Lemeshow pval	0.02				
Observations	773	773	773	773	773
Controls	Yes	Yes	Yes	Yes	Yes
<b>Nausea</b>					
Treatment	-0.01	-0.52	-0.01	-0.23	-0.01
	(0.01)	(0.50)	(0.01)	(0.20)	(0.01)
	0.29	0.30	0.34	0.25	0.28
R <sup>2</sup> / Hosmer-Lemeshow pval	0.00	0.00		0.00	
Observations	773	773	773	773	773
Controls	Yes	Yes	Yes	Yes	Yes
<b>Any symptom</b>					
Treatment	0.03	0.30	0.03	0.16	0.03
	(0.04)	(0.33)	(0.03)	(0.18)	(0.03)
	0.38	0.38	0.37	0.37	0.37
R <sup>2</sup> / Hosmer-Lemeshow pval	0.01	0.75		0.87	
Observations	774	774	774	774	774
Controls	Yes	Yes	Yes	Yes	Yes
<b>Too sick</b>					
Treatment	-0.02	-0.30	-0.02	-0.15	-0.02
	(0.02)	(0.29)	(0.02)	(0.15)	(0.02)
	0.30	0.31	0.31	0.30	0.30
R <sup>2</sup> / Hosmer-Lemeshow pval	0.01	0.93		0.76	
Observations	774	774	774	774	774
Controls	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, followed by pvalues

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix 8 (a): Ex-post minimum detectable effect size calculations**

VARIABLES	(1) Ex-Post MDE	(2) OLS Treatment Coefficient	(3) OLS Standard Error	(4) OLS Observations
Snack-test 1	0.198	-0.10	(0.08)	146
Snack-test 2	0.223	0.13	(0.09)	119
Both snack-tests	0.198	0.07	(0.08)	119

**Appendix 8 (b): Ex-post necessary sample size calculations for the observed effect sizes**

VARIABLES	(1) Ex-Post Necessary Sample Size	(2) OLS Treatment Coefficient	(3) Observed Standard Deviation	(4) Proportion Treated	(5) OLS Observations
Snack-test 1	595	-0.10	0.49	0.537	146
Snack-test 2	352	0.13	0.49	0.537	119
Both snack-tests	1164	0.07	0.48	0.537	119

**Appendix 9: Effect of toy-cheats on household member health  
(OLS results using an indicator variable)**

VARIABLES	(1) Health	(2) Diarrhea	(3) Flu	(4) Nausea	(5) Any	(6) Too sick	(7) Sick days
Treatment & NOT a toy cheat	0.06 (0.17) 0.71	-0.01 (0.02) 0.58	0.02 (0.04) 0.66	-0.02* (0.01) 0.07	0.00 (0.04) 0.97	-0.02 (0.03) 0.42	-0.05 (0.15) 0.75
Treatment & IS a toy cheat	-0.01 (0.15) 0.96	0.01 (0.02) 0.79	0.07* (0.04) 0.09	-0.00 (0.01) 0.95	0.07 (0.04) 0.12	-0.03 (0.03) 0.31	-0.16 (0.14) 0.23
Observations	768	769	769	769	770	770	768
R-squared	0.08	0.00	0.03	0.01	0.02	0.01	0.01
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses, followed by pvalues

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1