ECONSTOR

WWW.ECONSTOR.EU

Der Open-Access-Publikationsserver der ZBW – Leibniz-Informationszentrum Wirtschaft
The Open Access Publication Server of the ZBW – Leibniz Information Centre for Economics

Rieckmann, Johannes

Conference Paper

Determinants of drinking water treatment and hygiene habits in provincial towns in Yemen

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Environmental Economics III, No. C13-V3

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Rieckmann, Johannes (2015) : Determinants of drinking water treatment and hygiene habits in provincial towns in Yemen, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Environmental Economics III, No. C13-V3

This Version is available at:
http://hdl.handle.net/10419/113183

Nutzungsbedingungen:
Die ZBW räumt Ihnen als Nutzerin/Nutzer das unentgeltliche, räumlich unbeschränkte und zeitlich auf die Dauer des Schutzrechts beschränkte einfache Recht ein, das ausgewählte Werk im Rahmen der unter
→ http://www.econstor.eu/dspace/Nutzungsbedingungen
nachzulesenden vollständigen Nutzungsbedingungen zu vervielfältigen, mit denen die Nutzerin/der Nutzer sich durch die erste Nutzung einverstanden erklärt.

Terms of use:
The ZBW grants you, the user, the non-exclusive right to use the selected work free of charge, territorially unrestricted and within the time limit of the term of the property rights according to the terms specified at
→ http://www.econstor.eu/dspace/Nutzungsbedingungen
By the first use of the selected work the user agrees and declares to comply with these terms of use.

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Mitglied der Leibniz-Gemeinschaft

Determinants of drinking water treatment and hygiene habits in provincial towns in Yemen

Abstract

Development projects focusing mainly on the supply side of infrastructure repeatedly turn out to yield poor impact when being rigorously evaluated. This is also the case for an intervention connecting urban households in Yemen to piped water and improved sewerage networks. This study investigates the determinants of drinking water handling and hygiene behaviour. It aims at helping to improve retroactively outcomes, and to avoid pitfalls in the design of future projects. I use multivariate regression to identify the drivers of drinking water treatment and four hygiene habit measures. Connection of households is one of those drivers; however not the most influential one. Conducive water handling and hygiene behaviour appear to be responsive especially to training, access to information and communication technology (ICT), and school education. This paper contributes to the literature by a case study showing the importance of demand side aspects of infrastructure provision. Pipe and sewer grid construction must be flanked with enhancing point-of-use maintenance – or restoration – of drinking water safety. Future impact evaluations of programmes promoting good water handling and hygiene practices will be particularly useful when intervention assignments can be randomized, and baseline surveys provided for.

Keywords: ICT, Hygiene, Rigorous Impact Evaluation, Sewerage, Training, Urban, Water Handling, Water Supply, Yemen

JEL classification: D19, I10, I38, Q53

1. Introduction

"The General Assembly […] 1. Recognizes the right to safe and clean drinking water and sanitation as a human right that is essential for the full enjoyment of life and all human rights; […]."

General Assembly of the United Nations[1]

Development cooperation projects focusing mainly on the supply side of water and sanitation infrastructure repeatedly turn out to yield limited to poor impact when being rigorously evaluated. Zwane and Kremer (2007), for example, study the limited effect of community-level rural water infrastructure on diarrheal diseases. The World Bank (2008) indicates the prevalent absence of noteworthy "health gains for water treatment at the source". The intended impact on beneficiaries is often either not achieved at all, or diluted. In spite of considerable allocation of financial and human resources during planning and implementation, the benefits often seem to vanish unaccountably.

In order to fathom the underlying reasons and cope with this challenge, economists consider behavioural aspects in their research for several years. Waddington and Snilstveit (2009) emphasize that "interventions are embedded in social systems that have a strong bearing on their uptake and impact in the real world". Such social systems comprise belief and value systems, socio-economic factors, experience and behaviour. While infrastructure project designers typically consider sustainability aspects of supply[2] during planning, monitoring and when evaluating completed projects, it is a more rare practice additionally to pay attention to behavioural demand side aspects. For water and sanitation interventions, aspects of water handling and hygiene practices within the treatment beneficiary households are such demand side aspects of particular concern.

[1] Quoted from Resolution 64/292 (United Nations General Assembly, 2010), passed 28 July 2010, 108th plenary meeting.

[2] Donors today typically take into consideration malversation of financial resources as a possible cause of project inefficiency (see, for example, Brett, 2003, and Ebrahim, 2003). However, medium- and long-term technical aspects do not always receive adequate attention. Conceptualizing piped drinking water networks in water-scarce locations, for example, requires taking into account expected population growth due to demography and migration; as well as groundwater depletion and replenishment, and competing usages (*e.g.* for agriculture and industrial application). A lack of groundwater supply, for instance, might necessitate temporary rationing and partial disconnect of the piped network in order to maintain operational water pressure. This, in turn, may induce deterioration of water quality on its way to the household due to intrusion of wastewater into underused pipes through couplings and leaks; and growth of biofilm on their inner surface (see Barry and Hughes, 2008, and Flemming, 2009). Furthermore such intermittent drinking water supply requires households to either stock water in storage tanks filled during times of supply, or acquire water from potentially unregulated providers as truck water vendors. For a more detailed study of the effect of water rationing and permeable network pipes, and water storage and handling at household level, on drinking water pollution at the point of use see Lechtenfeld (2012).

This applies also to an intervention connecting households in provincial towns in Yemen to piped water supply and improved[3] sewerage networks (see Klasen *et al.*, 2011, for a detailed impact evaluation report). The impact on health not only falls short of ex-ante expectations. Klasen *et al.* (2012) estimate the intervention even to be detrimental for a large share of the beneficiary households. Only part of this change for the worse seems to be rooted in supply-side factors as intermittent water provision; and bacteriological contamination in the feed pipes occurring already before these reach the households. Water quality tests conducted (*ibid.*) reveal pollution with e.coli germs increasing between point of entry and point of use in at least one out of four households. Intra-household (and thus demand side) factors hamper the potential intervention impact. Lechtenfeld (2012) finds that particularly water storage and handling practices as well as personal hygiene habits have a share in worsening health outcomes.

The line of argument motivating the study at hand is straightforward: aspects of welfare and development as income, education and general wellbeing are fostered by good health. Health, in turn, is fostered by safe drinking water handling and good hygiene practices. These behavioural facets are subject to drivers on environmental, household and individual level. Identifying and quantifying these allows targeted policy designing.[4] The literature establishing the relevance of clean drinking water and hygienic sanitation is extensive, revealing its impact on health; and the dire societal consequences of diseases caused by polluted drinking water and unhygienic sanitation.

The mechanism through which water handling and hygiene practices affect the health of household members works directly: these practices affect occurrence and intensity of exposure to infectious waste and waterborne pathogenic germs. Waddington and Snilstveit (2009) show an illustration of the transmission routes of faecal-oral contamination through “fingers, flies, fields, food, and unclean waters”, along which germs are transported, proliferate and cause diarrhoea, dysentery and other water-related diseases after entering the body. It is widely recognised that contamination of drinking water by contact with hands contaminated with faeces adversely affects its quality at point of use independently from pollution at the source (see, *e.g.*, Jensen *et al.*, 2002). Trevett *et al.* (2005) show that distinct “evidence of disease transmission from re-contaminated drinking water exists”; and particularly link water storage and handling with incidence of diarrhoea.

[3] The term improved sewerage is used for sewer pipes transporting wastewater and faeces away from households to a communal wastewater treatment plant. This system is widely regarded to be preferable to traditional systems as open canals, open or covered cesspits, or septic tanks in vicinity of the house.

[4] It also is a prerequisite for (re-)establishing the effectiveness of the evaluated Yemenite intervention.

The relevance of diseases related to drinking water and insanitation for human development can hardly be overrated. Rosen and Vincent (1999) report that "poor water supply, sanitation and hygiene" account for 10.7 per cent[5] of mortality and 10.1 per cent of disability-adjusted life years (DALY)[6] lost in the preceding decades. They state that nearly all of these deaths and approximately 85 per cent of the loss in quality and time of life are caused by diarrhoeal diseases, the overwhelming majority of the affected being children.[7] Walker *et al.* (2013) estimate that diarrhoea caused 700,000 fatalities in 2011, with 72 per cent of these concerning children less than two years of age. Berkman *et al.* (2002) deem diarrhoeal diseases occurring in early childhood to impede significantly the development of cognitive abilities, which influences educational attainment and productivity. They cause direct cost of medical care (see for example Patel *et al.*, 2013, for the case of Indian slums) and indirect cost by foregone income due to DALY lost (see Lange *et al.*, 2012, for the case of South Africa). Diarrhoeal diseases decrease furthermore general quality of life and well-being (*e.g.*, see Prüss-Üstün and Corvalán, 2006).

Down-to-earth protective measures suitable to limit the risk of such diseases are implementable in Yemen. Households have to rely on water storage of some kind in absence of continuous on-demand supply with drinking water.[8] In arid urban settings, this is in most cases a roof- or courtyard-sited water tank of considerable capacity.[9] Otherwise, households rely on smaller plastic jerry cans, open buckets or traditional clay jugs. Besides some common-sense precautions regarding the storage of the water (*i.e.*, routine cleaning of the tanks, use of sealing lids, keeping containers and tanks out of reach of children and animals, and cooling[10]), it is of importance to additionally treat it before drinking. In order to assure the potability of the water, it can be dealt with in several ways. Neutralization of potential contamination with bacteria from the particularly dangerous e.coli[11]

[5] The numbers reported refer to Africa and shall serve as indication of the relevance and impact of diarrhoeal diseases, rather than allow direct conclusions about disease incidence and consequences in the Middle East.

[6] Disability-adjusted life years are a measure developed by the World Bank and the World Health Organization (Murray, 1994) in order to be able to compare the burden of different diseases. It takes into account the constraints an affected person is subject to during daily life, *i.e.* the severity of the disease.

[7] While the World Health Organization observes a distinct decline of child diarrhoeal mortality in 2003 compared to 1990 "from 159 to 70 deaths per 1000 live births", it estimates that still 21 per cent of deaths of children younger than five years are caused by diarrhoeal diseases in developing countries (Kosek *et al.*, 2003). Diarrhoea continues to be one of the leading causes of child death; with now unimproved mortality rates between 2005 and 2010 (see Santosham *et al.*, 2010).

[8] Such continuous on-demand supply is absent when households procure water from truck water vendors, haul it from a well by donkey cart tanks or handheld containers or have to bridge temporal disruptions of piped water supply.

[9] In Yemen, this capacity typically ranges between two and three cubic meters.

[10] Most storage tanks are steel roof tanks; and thus directly exposed to sunlight during most of the day.

[11] Germs from the family of *Escherichia coli* are responsible for annually "at least 120 million cases" (Feil, 2012) of bloody diarrhoea (dysentery), with their pathovar *Shigella* alone causing about 1.1 million fatalities (60 per

family, can for instance be achieved by chlorination, filtration or boiling of the water.[12] Hand washing with soap as well as water treatment reduces incidence and severity of diarrhoea considerably. The toilet being in a hygienic state decreases the risk of contamination of hands with faeces. There is "overwhelming evidence that hand washing, sanitation, and household and point-of-use water treatment improve health outcomes" (World Bank, 2008). Ejemot-Nwadiaro *et al.* (2012) find that promotion of hand washing reduces diarrhoea incidence in children by 32 per cent in low- or middle-income countries. By reducing the presence of bacteria of potential faecal origin more effectively[13] than hand washing using water without additives, the application of soap decreases the exposure of the human immune system to precarious germs. The soap improves the mechanical removal effect of the water and adds a chemical germ-killer capacity.

External validity is most likely. Yemen is one of the MENA (Middle East and North Africa) countries suffering from high water stress,[14] displaying a high rate of population growth,[15] high inequality,[16] and low education rates.[17] It is severely affected by the political unrests and social upheavals associated with the "Arab Spring"[18] which by today has affected more than a dozen countries perceptibly. I assume that the set of influences shaping intra-household behaviour is similar to that in

---

cent of which are children under five years of age) in developing countries (Niyogi, 2005); and provoke severe cramps. Presence of e.coli in water is a sign of recent contact with traces of human or animal faeces.

[12] See also the technical notes in Appendix 1 for further details on water treatment methods.

[13] Burton *et al.* (2011) observe in England a reduction of 21 percentage points using water alone and 36 percentage points using off-the-shelf soap (special antibacterial soap is yet more effective, but may not always be commercially available in developing countries). Cairncross *et al.* (2010) in their extensive meta-analysis estimate the reduction of risk to suffer from diarrhoea to reach 48 per cent. Luby *et al.* (2006) report a prevalence of diarrhoea 51 per cent lower in Pakistani neighbourhoods where hand washing with soap is promoted, compared to the control group. Aiello *et al.* (2008) find in a meta-study drawing on publications from 1960 to 2007 that improved hand hygiene reduces gastrointestinal illness by 31 per cent, the most effective improvement being "hand hygiene education with use of nonantibacterial soap".

[14] The country faces "formidable freshwater management challenges" (Haidera *et al.*, 2011) and a groundwater overdraft of twice the rate of recharge by rainfall (Taher *et al.*, 2012) – with some aquifers supplying urban population agglomerations depleting even faster.

[15] Figures differ between several official sources. According to the Central Intelligence Agency of the United States of America (2013) Yemen ranks place 30 out of 232 countries considered, with an annual average population rate of 2.5 per cent. The United Nations estimate 3.02 per cent (2012). The World Bank (2013a) estimates 3.06 per cent, corresponding to rank eleven out of 245. In urban areas the rate is even higher at on average 4.78 per cent (United Nations, 2012), which corresponds to a doubling of the urban population within less than 15 years. Zeug and Eckert (2010) estimate the annual population growth rate for the capital Sana'a to even have reached 7.3 per cent in 2007. The fast surge of the urban population aggravates the under-supply with clean drinking water.

[16] The Yemenite population is characterized by high inequality with a Gini coefficient of 0.377 (World Bank, 2005).

[17] The World Bank World Bank (2013b) indicates an average 3.7 years of schooling.

[18] The "Arab Spring" movement got under way in late 2010. In Yemen, associated demonstrations gained momentum by mid of January 2011, thus placing the Yemenite society among the first to climb on the bandwagon of – mostly – civil disobedience and protest against the powers. The process of reshaping allocation of political power and institutions is on-going.

many other countries in the Arab language area, African countries adjacent to the Sahara desert, and arid Central Asia. Such similarity increases the external validity of the findings and positions the paper as potential input for influencing development cooperation policy in the mentioned regions.

The main contribution of this paper to the body of literature is the minuteness of the analysis regarding direction and size of influence of a set of household characteristics on several behavioural outcomes in the MENA region. Based on the findings I derive recommendations regarding potential toeholds for influencing and coaching household behaviour. The main findings refer to determinants assessable in the medium term; which thus are relevant for policy design. *Drinking Water Treatment* is positively correlated with hygiene training, connection to piped water and to improved sewerage, exposure and access to ICT, and formal education. Training, connection to piped water and ICT seem to be distinctly conducive to the bundle of *Hygiene Habit* regressands.

I structure the remainder of this paper is as follows: Section 2 reviews relevant literature and derives hypotheses to be tested. Section 3 provides brief overview of the programme background and data. Section 4 introduces the empirical strategy and methods employed. Section 5 presents the empirical results. Section 6 concludes with a general discussion of the findings.

2. Literature review and hypotheses on determinants of water treatment and hygiene practices

The relevant literature lists an array of potential determinants of drinking water treatment, hand washing and other hygiene habits. Based on these and own first-hand observations regarding the circumstances in Yemen, I develop several hypotheses. These relate to determinants which firstly influence the likelihood of a Yemenite provincial household treating its drinking water before consumption, and secondly the hygiene behaviour exercised in daily life at home.

In the following section, I derive the expected direction of effects for these determinants, along with its respective transmission channels. Both *Drinking Water Treatment* and *Hygiene Habits* are deemed to be influenced by characteristics of the household as a whole (hypotheses 1 to 4) and the head of the household in particular (hypothesis 5) as itemised below. I indicate *Drinking Water Treatment* by a single binary variable, while I capture *Hygiene Habits* by four different dependent variables.[19] The rationale for this structure is the separation of ingestion-related from environment-related factors.

[19] Hygiene habits are measured by the reported practices and resource allocation linked to personal hygiene and health-relevant aspects of housekeeping. The dependent variables capturing hygiene habits are: *Use of Cleanser* for and washing, *Spending on Soap*, *Spending on Cleanser* understood as a combination of soap and detergent, and *Frequency of Latrine-Cleaning*. These are described in detail in the data section of this chapter.

Often – but not always – I expect the coefficient of the respective household to have the same sign for the dependent variable *Drinking Water Treatment* on one hand and all four alternative dependent variables indicating *Hygiene Habits* on the other hand. I expect the correlation coefficients of these latter variables to be consistent in sign, independent of which one I examine. The estimated sizes might be different, though. The hypotheses identified by character *a)* bear on *Drinking Water Treatment*, while those identified by character *b)* relate to *Hygiene Habits*.

As the intervention in Yemen centres on *infrastructural factors*, these are of high interest for this study. Such factors, as connection to piped drinking water and improved sewerage, are frequently indicated in the literature to play a role. Schmidt *et al.* (2009) list water access as determinant of hygienic behaviour in Kenya, Devine (2010) goes a step further and suggests that "convenient access" to water and soap is a key factor. Sugita (2004) finds that in Uganda the amount of water available per day and capita is a more important factor shaping hand washing behaviour than other factors.

I expect a household to see no particular necessity to look further after drinking water that it receives via an improved pipe network system. As communal treatment plants chlorinate the water before feeding it into the system, households with continuous piped water supply should have less reason to worry about its potability than those that buy their water from a vendor. What is not perceived to be necessary tends not to be done; so I assume households connected to the piped water network to have a lower[20] probability of treating their drinking water frequently than non-connected households. I therefore derive on the role of piped water ***Hypothesis 1a:*** *connecting a household to piped water <u>decreases</u> its likelihood of treating its drinking water before consumption.*

Connection to piped water should reduce the effort – while not necessarily the expenditures – of procurement of water needed for hand washing and cleaning chores. The increased volume of water conveniently available at home reduces the need for parsimony. I expect a higher average volume of water to be disposable within the household for cleaning of bathrooms and kitchens, and for taking measures of personal hygiene. While I do not expect a particularly strong effect here, I state ***Hypothesis 1b:*** *connecting a household to piped water <u>increases</u> its likelihood of displaying hygiene habits.*

I expect the effect which connection to improved sewerage (conditional on connection to piped water) has on the likelihood of treating drinking water to have the same sign as mere connection to piped water. Connection to yet another element of improved infrastructure should further foster the

[20] If not mentioned otherwise, the comparisons in the hypotheses are "ceteris paribus", and I assume all other determinants to remain unchanged.

general feeling of improved health safety; and thus reduce the efforts undertaken by the household to take basic precautions. Anyway, as sewerage is located "behind" consumption, the link to drinking water quality before consumption is not as obvious as that of water supply. Therefore, I expect the effect to be smaller than that of connection to piped water from the former section. I state ***Hypothesis 2a:*** *additional connection of a household to improved sewerage (i.e., on top of connection to piped water) decreases its likelihood of treating its drinking water before consumption, but to a lesser extent than connection to piped water.*

I assume the same mechanism to be at work as described in Hypothesis 2a when it comes to hygiene habits. I expect the perception of reduced health hazard coming along with the improvement of infrastructure to reduce the efforts undertaken by household members to keep the bathroom clean; and to consistently maintain hygiene routines.[21] Therefore I state ***Hypothesis 2b:*** *additional connection of a household to improved sewerage decreases its likelihood of displaying hygiene habits, but to a lesser extent than connection to piped water.* Note that I expect the role of improved sewerage to be similar for water treatment, but to be converse for hygiene habits.

Hygiene training and awareness campaigns are consistently associated with favourable behavioural change. Curtis *et al.* (2011) state that "promotion of safe hygiene is the single most cost-effective means of preventing infectious disease" with a focus on hand washing and safe stool disposal being especially beneficial, and commends to promote hand washing and hygiene "at least as aggressively as vaccination". Luoto *et al.* (2011) estimate that the use of "persuasive social marketing messages that harness findings from behavioural economics" increase occurrence of water treatment,[22] while stating that sustainability of behavioural change may depend on repeated exposure.

I assume that exposure to hygiene trainings – as conducted by various governmental and non-governmental entities in Yemen – increases awareness about the link of contaminated water with water-related diseases, as well as knowledge about potential sources of water contamination within the household; and about feasible precautionary measures. I therefore state ***Hypothesis 3a:*** *those households in which at least one member participated in any type of hygiene training in the past have*

[21] The latter comprise hand washing following, *e.g.*, each emiction or defecation, washing infants and toddlers, and before preparing and eating food.

[22] The latter authors refer to Kenya. The studies are in line with a multitude of others: for example, Parker *et al.* (2006) show that training in household water chlorination as well as in six-step hand washing procedures lead to promising levels of sustainability in learned behaviour in rural Kenya. Galiani *et al.* (2012) support this by showing that a Peruvian campaign embedding schools increases the awareness of the population regarding the role of hand washing and soap use, and leads to application that is more frequent. Chase and Do (2012) find in a Vietnamese setting that such campaigns have the intended direction of effect, while large-scale campaigns still face challenges in terms of the size of the effect. Pengpid and Peltzer (2012) report hygiene campaigns addressing schoolchildren to be effective in African countries.

a higher likelihood of treating their drinking water before consumption. In line with the reasoning Hypothesis 3a is based on, I assume enhanced awareness and background knowledge to motivate household members to develop and maintain hygiene habits. Thus I state ***Hypothesis 3b:*** *those households in which at least one member participated in any type of hygiene training in the past have a higher likelihood of displaying hygiene habits.*

Exposure to information and communication technology – such as mobile phones and particularly mass media like television, radio broadcasting, the World Wide Web – may have an impact on water treatment and hygiene habits displayed by household members due to role model effects and access to information. The evidence is mixed so far. Galiani *et al.* (2012) indicate that a mass media and communications campaign on community level[23] has no effect on knowledge and behaviour concerning hand washing. In contrast to that, Schmidt *et al.* (2009) find that both exposure to and ownership of media are linked to soap use.

I assume that households that have access to radio or television broadcasting, the World Wide Web, or telecommunication devices are better informed about water-related diseases and feasible precautionary measures of water handling than households that are more isolated. Awareness campaigns undertaken by governmental and non-governmental entities probably address them more effectively. The household members can more easily gather information on their own initiative; and exchange information about best practises with more other households. On the role of ICT, I state ***Hypothesis 4a:*** *those households with exposure to information and communication technology have a higher likelihood of treating their drinking water before consumption.* Following the same line of argumentation I further state ***Hypothesis 4b:*** *those households with exposure to information and communication technology have a higher likelihood of displaying hygiene habits.*

Various other *household characteristics* are listed in the literature to be determinants. Aunger *et al.* (2010) mention the "habit of hand-washing at particular junctures during the day"[24] to constitute a significant predictor of hand-washing behaviour. The financial resources a household can dispose of may represent an economic constraint, which together with perception of impracticality of safe household practices can prevent households from using these practices. In Tanzania this is the case even when mothers are aware of the role of these practises in mitigating health risks (see Badowski

[23] The study refers to a campaign in provincial Peru.

[24] The authors refer to the case of Kenya. In a Yemenite context the daily Islamic prayer times may represent such junctures due to the ritual washing (*Wudu'*), but it has to be pointed out that no soap is used in this ritual.

et al., 2011). Begum *et al.* (2011) identify three features of adoption behaviour[25] concerning improved water and sewerage. These are namely wealth (the poor have a lower probability of benefiting from improved drinking water supply and sewerage), education (displaying a threshold effect with secondary or higher education increasing the probability), and household size (lowering the probability). The presence of children and perception of diseases as severe both increase compliance with hygiene guidelines (see Zomer *et al.*, 2013). Totouom *et al.* (2012) study the factors in Cameroon which determine if a household opts to treat its drinking water. They state that household level of education has a positive effect on the probability of water treatment. The same is reported about wealth, illness history (increasing the probability due to heightened risk-aversion) and the number of children younger than five years of age living in the household (with a positive effect on the probability due to increased sensitivity regarding water-related health issues). These characteristics on household level captured during the interviews have a conjoint (*i.e.*, on all or at least most of the household members) effect.

This is why I control for household composition, wealth and illness history as well as awareness regarding water-related illnesses. I assume that households with a large proportion of needy members are managed by adults who tend to be more cautious regarding water-related health issues than those in households where more members are self-reliant. Where several young children or elderly adults live, the occurrence of diarrhoeal diseases generates a direr situation than where household members are endowed with a more robust physical health. I expect households with a higher number of members who have to be looked after – *i.e.*, children and the elderly – by other members have a higher likelihood of treating their drinking water before consumption, and of displaying hygiene habits.

I further assume that wealthier households tend to be those which are populated by more educated adult members, which should influence their awareness about precautionary measures regarding health and water safety. Severe (water-related) diseases probably have been affecting them less, because they took these precautionary measures more often. Furthermore, they dispose of more economic resources providing them with additional options as, *e.g.*, acquisition of water filtration systems. I expect that wealthier households will have a higher likelihood of treating their drinking water before consumption and of displaying hygiene habits. The understanding is that supplies

[25] Note that in the context of this study in Yemen, households have no choice whether to be connected to improved piped water or improved sewerage. Still, one can reasonably argue that the adoption behaviour studied by Begum *et al.* (2011) for the case of Bangladesh shows similar characteristic traits as adoption behaviour of intra-household drinking water treatment.

acquired with additional available economic resources refer to soap, laundry detergent and household cleaning agents rather than hardware here.

Third, I assume that in a household that was exposed to water-related diseases within the recall period of four weeks before the interview household members[26] take precautionary measures regarding water safety. (Note that the opposite relationship is conceivable.[27]) A burnt child dreads the fire, but the effect might last for a limited time only when it suggests taking rather bothersome additional measures without immediate observable gain. Consistently, adults who experienced water-related diseases might be more cautious also regarding hygiene habits than others who have been luckier. I expect that households in which at least one member suffered from water-related diseases in the recent past have a higher likelihood of treating their drinking water before consumption, and displaying hygiene habits. (Again, I could reverse the argument.[28])

Individual *characteristics of the household head* affect the probability of drinking water treatment and hygiene habits. The relevant literature provides mixed evidence regarding the role played by gender of household headship, while age and education[29] seem to have a positive effect. Tan Jr. and Capuno (2012) estimate the effect of the mother or father holding at least a high school degree on probability of drinking water treatment to be positive, the age of the mother having a positive effect, and a positive effect for the gender of the households head being male. Sattar *et al.* (2007) show that in Pakistan "the education level of female decision-makers compared to that of male decision-makers has more significant effect in using any […] method of purification". They argue that women are more prone to adopt water purification practices "because females are in general more intensively involved in the food related household activities". Totouom *et al.* (2012) on the other hand indicate that male household headship increases the probability of drinking water treatment, but point out that this is in contradiction with other studies. Begum *et al.* (2011) find that the gender of the household head is not significantly correlated with adoption behaviour concerning improved water and sewerage.

---

[26] I here additionally assume that those members who are tasked with household chores of water procurement and storage know about the link of contaminated drinking water with these diseases.

[27] Of course, it might be the case that an opposing relationship is at play here as well: recent symptom occurrence might reflect habitual disregard of precautionary measures, in contrast to the former hypothesis reflecting a long-term effect.

[28] Habitual disregard of precautionary measures might have contributed to the occurrence of symptoms, and persist.

[29] Both, duration and level of education appear to play a fostering role. Sugita (2004) finds that duration of education increases the probability of drinking water treatment, together with awareness about diarrheal risk factors and attitudes regarding hand washing. Schmidt *et al.* (2009) identify the level of education to be important.

I assume that over the period of formal education – in school, vocational training, and university – knowledge is taken up on links of water contamination with water-related diseases, safe water handling, and hygiene behaviour. This knowledge partially translates into formation of habits and implementation of best practices. Based on this, I state ***Hypothesis 5a:*** *the more educated the head of the household is the higher is the likelihood of the household treating its drinking water before consumption.* Again, the reasoning for this hypothesis is the same as for the next. I consistently state ***Hypothesis 5b:*** *the more educated the head of the household is the higher is the likelihood of the household members displaying hygiene habits.*

In Yemen mainly the women look after needy household members, and spend more time at home than their husbands. I therefore assume that female household heads tend to be more aware of health issues among the members; and can better safeguard adequate hygienic and water handling behaviour. As previous literature hints towards its influential role, I control for gender. I expect households with female headship to have a higher likelihood of treating their drinking water before consumption, and of displaying hygiene habits.

The additional treatment of drinking water following its procurement is neither common nor traditional in Yemen. Regarding the role of age, I therefore argue opposed to the literature mentioned above, and control for it in the regressions. I assume that those cohorts with more recent exposure to formal education have taken up habitual water treatment with higher probability. Promotion of best practices probably influenced their formation of opinion and habits within the previous three decades. I further assume hygiene habits to have been internalised more intensely by the younger generation. I expect the head of the household being younger increases the likelihood of treating the drinking water before consumption, and of displaying hygiene habits.

The study on hand tests the aforementioned hypotheses, investigating the conjectured determinants of the probability of drinking water treatment and habitual exercise of hygiene practises. I empirically assess – first for water treatment, then for the bundle of four dependent variables available to proxy hygiene habits – each hypothesis, using Yemenite household survey data. The empirical analysis reveals evidence supporting some of these hypotheses. Others I have to reject; for a few of them, I do not find insightful evidence.

3. Programme background and data

While housing one of the oldest civilizations on earth, the Republic of Yemen is a young state. The country had been separated into the Yemen Arab Republic (called North Yemen) and the People's Democratic Republic of Yemen (called South Yemen). Its reunification into the Republic of Yemen in its current borders became effectual in 1990. While infrastructure in the whole country is

developable compared to other countries located on the Arab peninsula, the North is lagging behind the South for historical reasons.[30] One of the committed bilateral donors with whom the current Yemenite government cooperates is the Federal Republic of Germany, represented by its Ministry of Economic Cooperation and Development (BMZ).[31] Besides basic education, the main foci of the development cooperation are drinking water supply and wastewater disposal. In the year 1989 the Provincial Towns Program (PTOP) was initiated in eight Yemenite towns, all of which are located in the area of the former North Yemen.[32] In these locations, existing piped water and sewerage networks – including water works and wastewater treatment plants – were rehabilitated, extended or fundamentally constructed.

The German financial contribution was transacted by the development banking branch of the *Kreditanstalt für Wiederaufbau* (KfW), which subsequently commissioned a rigorous impact evaluation of two sample projects of PTOP in 2008. The data that this study is based on stem from this evaluation (see Klasen et al., 2011, for further details). KfW and the evaluation team of the University of Göttingen jointly selected two out of the eight provincial towns.[33] In order to capture contingent systematic differences between topographic regions, one project town – Amran – was chosen from the central mountainous region, the other one – Zabid – from the western coastal plain. The researchers identified suitable control towns located over the same water aquifer as the respective project town and in spatial vicinity (no more than 20 kilometres away)[34] based on topography, population size and socio-economic similarity. Further selection criteria are the existence of baseline data and variance of water availability within the survey town. The latter

[30] From the retreat of the Ottoman Empire in 1918 to the death of Imam Ahmad bin Yahya in 1962 the North was ruled in form of an autocratic imamate. Technological innovation stagnated for ideological reasons, and this part of the country became largely isolated from the international scene. Followed by the Northern Yemen Civil War and the revolution that was its result, the North became the more secular Yemen Arab Republic. It commenced to cooperate closely with its main ally Kingdom of Saudi Arabia. The Federal Republic of Germany recognized the Yemen Arab Republic as independent state right away in 1962 when the revolution began. It has been maintaining development cooperation since about 44 years. Around the same time, the Aden Emergency initiated the end of the British colonial rule in the South. The socialist state of South Yemen emerged in late 1970, quickly building close ties to the German Democratic Republic, the People's Republic of China, and the Soviet Union.

[31] The relevant development cooperation endeavours which this study is referring to are identified by BMZ reference numbers 1989 66 160, 1998 66 112, 2001 65 787, and 2002 65 165.

[32] These towns comprise besides Amran (located in the mountainous part of the country) and Zabid (located in the coastal plain *"Tihama"*) the urban agglomerations of Al Mansuriyya, Al Mukha, Bait al Faqih, Bajil, Hajja, and Yarim.

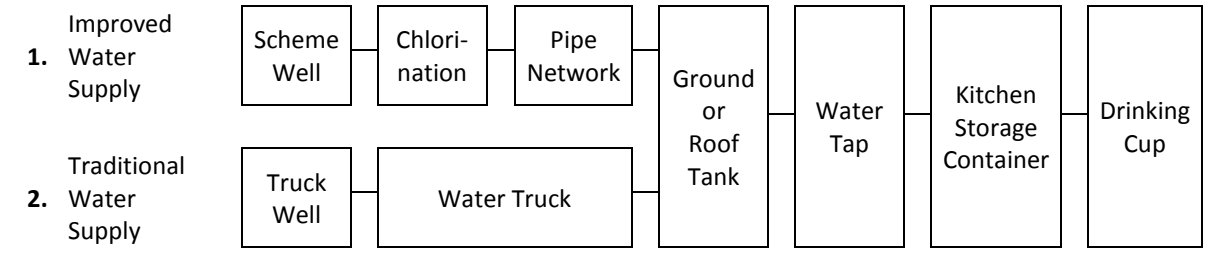
[33] For a detailed description of the Provincial Town Programme and especially the impact evaluation see Klasen *et al.* (2011).

[34] For the mountainous project town Amran the control town is Raydah, for the coastal project town Zabid it is Al Jarrahi.

provides the opportunity to construct in-town control groups as additional counterfactuals[35] besides the control towns. For the mountainous region, the control town is Raydah; for the coastal plain it is Al Jarrahi. Figure A.1 in Appendix 7 depicts the locations where the survey yielding the data for this study was conducted.

For an illustration of the technical scheme of water supply from well to point of use refer to Figure 1. Where water supply is intermittent – which is the case in the mountainous and to far lesser extent in the coastal project town[36] – steel storage tanks of average filling capacities ranging between two and three cubic metres are present in literally every affected household. In the control towns without piped water, nearly every household relies on such a storage tank; and periodically purchases drinking water from tanker trucks.[37] In times of availability of water, they fill the tank. The water gets pumped up[38] from street level – either fed from the piped water network, or from a water truck – into the storage tank. From there it typically runs through an intra-household pipe system to a kitchen water tap. Having cooled down during the night, the water usually is filled in the morning into a container holding up to 20 litres. Here it is retained for purpose of drinking and meal preparation during the day. These containers are commonly wide-necked pitchers made of porous clay to benefit from the evaporation chill either, or enclosed plastic jerry cans that ideally are equipped with spigots. Several household members often consecutively share the mugs into which they fill the water for drinking.

Figure 1: Water supply chain and test points



[35] In-town control groups are used in the main impact evaluation. In the study at hand they are used for descriptive statistics, but employment regarding the analysis is abstained from on the grounds of econometric sample size requirements. Subsample sizes are too small to render meaningful estimation results for the dependent variables of choice.

[36] In the coastal project town water supply is more reliable than in the mountainous area because ground water depletion is not as severe, and population growth due to demographics and migration is more moderate.

[37] Traditional water fetching from wells in smaller containers like buckets and jerry cans plays much less of a role in terms of frequency and water volume acquired. It leads to the same phenomenon in the household: drinking water is stored for some time before consumption.

[38] The storage tank is mostly located on the rooftop in order to be independent of electricity. Power blackouts are commonplace in Yemenite towns, during which diesel pumps would have to be used for distribution of the water within the household. Mounting the tank aloft allows exploiting gravity instead, a principle familiar from water towers located on high ground. Few households mount the storage tank in the courtyard.

The offside location of the storage tanks, combined with the default mounting of a sealing lid, makes contamination of the stored drinking water by household members or animals unlikely. Still, contamination of the water elsewhere within the household can occur due to unsafe water handling practices in combination with presence of germs due to a lack of hygiene. Critical junctures are the process of filling the water from the tap into the kitchen storage container, or directly into food; and from the kitchen storage container into the drinking mug or food. Germs adhering to hands, worktops, rags and inner surfaces of bowls, kettles *etc.* may enter the water and proliferate.

A comprehensive questionnaire-based[39] household survey, conducted in both treatment and both control towns in 2009, is the main data acquisition component[40] of the impact evaluation. Sampling was randomized, based on satellite image mapping with identification marking of rooftops. The team divided each town into square-shaped clusters. In order to cover all neighbourhoods, all of these clusters entered the sampling frame. From each cluster, a sub-sample was drawn using a random starting point and constant house-roof selection intervals. Pre-formulated consistent selection rules and replacement procedures applied. This was a precaution in case enumerators encountered more than one household in a single building, the house was not a residential building, or the enumerators could conduct no interview for other reasons.[41] Table 1 below displays the detailed composition of the sample of 2403 households used[42] in this study, by connection status and location.

Self-selection into the treatment – *i.e.*, connection to piped water and additional connection[43] to improved sewerage – is not likely. The reason is that the network was constructed starting from the water works[44] with large feed pipes leading into the inner town, then fanning out into smaller distribution pipes, and then streetwise into household supply pipes. The network was systematically extended from the inner town outwards, following the streets and hampered only by particular rocky areas where the digging of ditches – which partially was done manually due to an employment-

[39] The questionnaire used in the household survey is available in English language in the appendices of Klasen *et al.* (2011).

[40] The main household survey was supplemented by interviews based on short questionnaires conducted with commercial water truck well owners, water truck drivers, a subsample of households visited for water quality tests, pollution data from water quality tests, secondary data on water-related diseases collected from health facilities, school attendance and educational attainment data from education facilities, and secondary census and survey data from the Yemenite Central Statistical Organization (CSO). The chapter at hand resorts entirely to data from the main household survey.

[41] Most residential houses in the four provincial towns are inhabited by only one household. Interview refusal rates are close to zero and negligible.

[42] The survey covered 2518 households. The sample used in this study is reduced by 115 households to avoid missing observations.

[43] Connection to improved sewerage is conditional on connection to the piped water network.

[44] The water works chlorinate the drinking water extracted from five (mountains) respectively three (coastal plain) source wells at water treatment plants before feeding it into the distribution pipe network.

creation scheme – was particularly difficult. If a household in the evaluation sample is located in a street that today is connected to the grid, and was located at the same place at the time of the treatment,[45] then it de facto could not opt out of the treatment. Therefore, this study benefits from a natural experiment setup.

Table 1: Household sample, by region and connection status

Region	Connection status	Households	Percentage (local)	Percentage (global)
Mountainous	Water	199	17.61	8.28
	Sanitation	269	23.81	11.19
	Control	662	58.58	27.55
	Subtotal	1130	100.00	47.02
Coastal	Water	127	9.98	5.29
	Sanitation	714	56.09	29.71
	Control	432	33.94	17.98
	Subtotal	1273	100.00	52.98
Total		2403		100.00

To measure water handling and hygiene habits, I use five dependent variables. For an overview of mean outcomes by connection status, see Table 2 in Appendix 3. *Drinking Water Treatment* is a binary indicator variable taking the value 1 if a household subjects its drinking water to any kind of disinfecting treatment before consumption, and 0 otherwise. Self-reporting on hygienic conditions, cleanliness and related habits is prone to be subject to reporting bias. This is a special case of the *Hawthorne effect* (cf. Wickström and Bendix, 2000).[46] I therefore use four alternative measures in order to compose a more robust picture as close to reality as possible with the data at hand.

Use of Cleanser for Hand Washing is a binary indicator variable taking the value 1 if a household uses cleanser – this can be soap, laundry detergent as an alternative for soap, or a combination of these[47] – for hand washing, and 0 otherwise. I deem the variable to indicate habitual hygiene better than simple reporting of hand washing, as almost every respondent – those answering falsely probably

[45] Very few new houses were constructed after the treatment, apart from those in the outskirts of the towns. The reason is found in the high building density in the organically evolved old centres. Another possibility of self-selecting into the treatment group would be moving into existing houses (and replacing other households).

[46] The respondent is aware of the thoughtfulness and possible judgement of the interviewer; and possibly wishes to comply with what she thinks the wishes of the researchers are. This would be a manifestation of the Hawthorne effect, which is driven by behavioural change of the observed due to the observation. There is little reason to assume the related *John Henry* effect (Saretsky, 1972) to be at work, as the latter is concerned with behavioural change of the control group only, which might be motivated to outperform – or report to do so – the treatment group. The questionnaire design does not stimulate the respondent in any way to compete with other respondents belonging to households with a different connection status.

[47] Cleanser is defined in this study to be a combination of soap and laundry detergent. In Yemen as elsewhere in the Middle East and North Africa it is not unusual to replace soap with laundry detergent for hand washing. For this reason I consider both.

doing so craving for admiration or out of embarrassment – states that household members wash hands. *Soap Spending* is a continuous variable that gives the logarithmised expenditure per capita in Yemenite rial (YER) for soap, in a typical[48] month. The variable is logarithmised in order to approximate its distribution to a normal distribution. It is intended to render an alternative proxy for the level of hygiene within a household and is more finely gradated[49] then a binary variable. I drop observations[50] with an initial value larger than 500 YER (approximate 2.33 United States Dollars[51]) to avoid outliers which are suspected to report business-related expenditures, *e.g.* for laundry cleaner firms operating in the same house as the household is located in. *Cleanser Spending* is the corresponding logarithmised continuous variable for the combined expenditures of soap and laundry detergent. It is introduced serving as a robustness check. *Latrine-Cleaning Frequency* is a continuous variable indicating how many times per week the latrine of a household is reportedly cleaned. Like the two expenditure variables, it serves as a proxy for the level of hygiene maintained in a household. I conceive the variable as a proxy for general cleanliness efforts of the household. Its informative value might be somewhat restricted, as a positive reporting bias is likely. My assumption is that this reporting bias is not systematically different from household to household. Most households report to clean once or several times a week, some report to do so up to twice a day. Very few households report either never to clean or to do it more often than twice a day.

The main explanatory variables of interest indicate the connection status of a given household regarding piped drinking water, and additional connection to improved sewerage. I am particularly interested in the role of these variables as potential determinants of behaviour because the financial, organizational and technical input to the infrastructure programme is one of the main donor contributions in the field of public health in Yemen. A cost-benefit analysis needs to take into consideration direct as well as indirect effects of connection of households. It is an important question if an extensive intervention like the one at hand has a bearing on behavioural issues regarding handling of drinking water and other hygiene-related practices. *Piped Water* is a binary indicator variable taking the value 1 if a household is connected to the piped water network, and 0 otherwise. *Sewerage* is the corresponding binary indicator variable regarding additional connection to improved sewerage.

---

[48] Note that the Islamic month of fasting, *Ramadan*, around which consumption patterns vary from the rest of the year, is not a typical month. The survey and questionnaires are designed considering that.

[49] Note that I do not assume a linear relationship between spending and the level of hygiene exercised within a household. Diminishing marginal returns appear likely.

[50] Approximately the 1.6 per cent highest amounts are excluded for spending on soap, and 1.45 per cent for the combination of soap and laundry detergent.)

[51] Information on the currency exchange rate stems from Bartram *et al.* (2005).

I include other covariates in the analysis to control for characteristics of households as a whole, household heads as influential individuals, location, and media exposure. I try to employ housing characteristics as instruments, but these turn out to be of limited suitability.[52] The choice of this set of control variables is motivated by the review of existing literature, my hypotheses regarding possible determinants of water treatment and hygiene habits, and data availability. I cover characteristics of the household head by incorporating the *Age*, *Gender* and *Marital Status* as well as alternative measures of *Education* (level achieved, duration, and literacy). Household characteristics include alternative measures of composition (*Dependency Ratio,*[53] *Number of Household Members*, *Number of Children*, and *Number of Children and Elderly*), the *Mean Duration of Education* of adult members, water-related *Disease Awareness*, *Occurrence of Disease Symptoms* within a recall period of thirty days, and exposure to *Hygiene Training* in the past. I include a proxy for household wealth to control for unobserved intra-household infrastructure and hygiene practices. For this purpose, I use an *Asset Index* based on building material (roof, walls, and windows), ownership of the house, of a motor vehicle, and of particularly valuable electronic assets. It is calculated for the mountainous and the coastal region separately to account for regional differences. For details see Klasen *et al.* (2011). As an alternative to the asset index, I use the *Building Material* of the outer walls of the house.[54] I also control for presence of a *Water Storage Tank*, as a specific item of intra-household water supply hardware. *Location* effects are controlled for by including binary location indicator variables on region or town level. I control for *Exposure to ICT* and mass media including several alternative or supplementary variables (possession of a television set, personal computer, satellite dish, radio broadcast receiver, portable or landline phone, and access to the World Wide Web). Housing characteristics include the *Distance* of the house from the historic centre of the town, its reported *Age* and an indicator variable denoting whether it rests upon shapeable or *Rocky* subgrade. For a detailed description of all covariates, see Table 1 in Appendix 2.

4. Methodology and empirical strategy

I estimate the roles that the conceivable determinants play for the likelihood that a household resorts to a) *Treatment of Drinking Water* before consumption and to b) *Use of Cleanser* for hand washing using Probit, and four different instrumental variable estimation procedures, respectively.

[52] For the purpose of documentation (and possible future extension using better instruments) I still document the approaches tested and its findings in Appendices 4 and 6.

[53] The *Dependency Ratio* is defined as the relation of household members unable to contribute to household income due to their age (enumerator) to those in working age (denominator), regardless of actual employment status. The number of household members aged below 15 or above 60 is divided by the number of household members aged between 15 and 60.

[54] Estimation results for specifications including *Building Material* as covariate and alternative to *Asset Index* are not included in the results tables due to space limitations. They are available on request.

The reason for this wide approach lies in the nature of the data. The dependent variables as well as the main explanatory variables are binary, which necessitates the employment of several estimators to narrow down the results when trying to cope with potential endogeneity.[55] The three remaining continuous dependent variables scrutinized in this study – reported *Spending on Soap* or *Cleanser*, and the reported *Frequency of Latrine-Cleaning* – are regressed on the covariates using simpler methods. The basic regression model is based on Ordinary Least Squares (OLS). I apply the Two-Stage Least Squares (2SLS) estimator to control for endogeneity issues.

Endogeneity[56] regarding the main explanatory variable cannot be ruled out ex ante. Coping with it by instrumenting[57] the main explanatory variable requires methods which account for the fact that the two dependent variables as well as the explanatory variables to be instrumented are all binary. Reverse causality is not probable for either of the dependent variables, as households were connected based on their location within certain streets. Measurement error on the other hand might be possible due to reporting bias regarding explanatory variables (*e.g.*, falsely stating compliance with best hygiene practices).[58]

I use the same instrumental variables as Klasen *et al.* (2012). The respective – potentially endogenous – explanatory variable indicating the connection (either to piped water only, or additionally to improved sewerage) status is instrumented using three variables. These are 1) the distance of the house the household resides in from the centre of the town in metres[59] (*IV Distance*), 2) the age[60] of the house (*IV Age of House*), and 3) a binary variable indicating whether the house is located on rocky

[55] For further details refer to Appendix 4.

[56] Endogeneity is present in case the error term correlates with one or several of the covariates. The presence of endogeneity would make the reliable estimation of the influence of the various conceived determinants a difficult matter. It might arise when the dependent variable influences one or severely explanatory variables (reverse causality), due to measurement error or omitted variables. Omitted variables would have to be correlated with both the dependent as well as at least one explanatory variable in order to cause endogeneity. While there are no hints pointing towards an increased probability of either of three issues, none can be ruled out upfront. This study is not concerned with time series analysis, therefore the fourth possible cause of endogeneity – autoregression with autocorrelated errors – is not an issue.

[57] The matter can be addressed by instrumenting the covariate suspected to be endogenous. The instrumental variable or variables chosen must fulfil both the criterion of relevance (*i.e.*, actually have a non-negligible correlation with the covariate) and that of exogeneity (it or they must not correlate with the error term). The latter point can be read as the requirement that the instrumental variable may have an influence on the dependent variable only indirectly – through its effect on the possibly endogenous covariate – but not directly, controlled for the other variables in the model.

[58] Omitted variable bias cannot fully be excluded.

[59] A larger distance to the centre is expected to decrease the probability of a household to be connected to the pipe network, as the networks extend from the centres outwards.

[60] Older houses are usually located in the historic centres of the town, where the networks extend from outwards.

ground[61] (*IV Rocky Ground*). I test the chosen instruments regarding relevance and validity for each dependent variable. To test for relevance I compare the Cragg-Donald Wald F-statistic to the appropriate critical value, taking into account the number of instrumented variables, of instruments, and the "desired maximal size (r) of a 5% Wald test" (refer to table 5.2 in Stock and Yogo, 2005). To rest the assumption of exogeneity of the instruments beyond a pure argumentative line of defence, the Hansen's J statistic is calculated. I display details concerning these calculations in Appendix 4. While it turns out that I can be confident regarding the validity of the instruments, I cannot confidently show that the exogeneity assumption is fulfilled based on statistical tests. I therefore rely first and foremost on the basic Probit regressions. On basis of logical reasoning it is difficult to conceive of how the instrumental variables could not be endogenous. For purpose of documentation and facilitation of improving the analysis in case better instruments can be found, I provide further detail on the respective IV estimations (see Appendix 4) and their results (see Appendix 6). Still, those latter results have to be regarded with caution.

5. Results

I structure all sub-sections of this section in a way helping the reader keeping track of the various dependent variables and tests. To this end, I briefly recap my expectations. Then, at a time, I first show descriptives, then present the analysis including robustness checks, and third discuss the findings.

Connection to piped water and improved sewerage I expect reducing the diligence exercised when handling drinking water. Exposure to hygiene training or awareness campaigns should increase the probability of drinking water treatment. I expect the same direction of effect for wealth, education, number of children, and experience with water-related diseases, female headship, and age of the household head. The role of ICT and media exposure is unclear ex ante, while I tend towards assuming a positive effect. I do not form any expectation regarding the role of location, or marital status of the household head. The same pattern applies for all of the dependent variables intended to map hygiene habits other than probability of drinking water treatment. The expected sign of connection to piped water constitutes the single exception.

5.1 Drinking water treatment

Households in the Yemenite project towns tend not to treat their water in any way before drinking. In the mountainous region, only 16 per cent report to do so, while in the coastal plain, hardly

[61] Because the water and sewerage pipes are usually led along the streets subterraneous, and constructing ditches was often done without heavy machinery, rocky areas are assumed to have a lower probability of being covered by the pipe network.

anybody does (less than three per cent, see Table 2 below for further breakdown according to connection status). The underlying reason for the latter phenomenon might be that literally every household in the coastal project town is connected to piped water, the supply of which is in addition far less intermittent than in the mountain project town. Due to these circumstances water storage tanks are less prevalent, and trust in the quality of piped water might reduce the perceived need to treat the drinking water at home. Still, this cannot explain why drinking water treatment is so rare also in the coastal control town, where no piped water is provided. Only 221 out of 2406 households in the sample that I use for this part of the analysis report to treat water. Respondents report an array of methods; approximate relative frequencies are given in parentheses. These households filter (60%) or boil (28%) the water, let it stand and the contained particulate matter settle (7%), stain it through a piece of cloth (4%), add chlorine or bleach (0.5%), disinfect it by exposure of a transparent plastic container to sunlight or apply yet another method. It is eye-catching that the technically down-to-earth and cost-efficient disinfection method of chlorination is hardly used at household level. A possible reason for this is the unpleasant effect on taste when overdosing, but no information suitable to investigate this matter further is available in the data.

Table 2: Drinking water treatment, relative frequencies by location and connection status

Connection	Water Treatment	Mountain Project	Mountain Control	Total	Coastal Project	Coastal Control	Total
None	Yes	4.90%	8.08%	5.74%	0.00%	2.55%	0.86%
	No	39.00%	91.92%	52.87%	0.00%	97.45%	33.07%
Piped Water	Yes	4.43%	0.00%	3.27%	0.24%	0.00%	0.16%
	No	19.38%	0.00%	14.30%	14.86%	0.00%	9.82%
Sewerage	Yes	9.93%	0.00%	7.33%	2.73%	0.00%	1.81%
	No	22.37%	0.00%	16.50%	82.16%	0.00%	54.28%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
	N	836	297	1133	841	432	1273

Note: no households are connected to piped water or improved sewerage in the control towns. All households in the sample are connected at least to piped water in the coastal project town.

Eight major factors seem to determine the probability of a household in the sample treating its drinking water in any way before drinking. While the estimated size of the effects slightly varies across different specifications, the levels of statistical significance remain largely unchanged. The same holds for the relative order in terms of coefficient size.

Specifications (1) to (6)[62] in Table 3 below provide an overview, opposing the two covariates of main interest – *i.e.*, connection to *Piped Water* on the one hand and additional connection to *Improved Sewerage* on the other – and varying measures of education and location. I exhibit estimated average marginal effects. Complementary specifications are displayed in Table 9 to Table 12 in the results appendix; now split by main explanatory variables of interest (piped water only, and additional connection to sewerage).

For reasons of space limitation I respectively portion the tables into specifications varying ICT exposure variables on the one hand, and introduce location variables as well as alternative estimation methods on the other hand. The alternative estimators are intended to control for endogeneity, and second stages are displayed. The results of these latter specifications (see Table 11 and Table 12) as well as results of first stages (see Table 29) are displayed in Appendix 6. The coefficients of determination of the models range from 17 to 19 per cent.

The most influential covariates in terms of coefficient size are those indicating the location of the household. The household being located in the mountainous area[63] appears to increase the probability of treating its drinking water before consumption by 15 to 16 per cent, compared to location in the coastal plain. Other explanatory variables with the same positive sign of coefficients – always statistically significant at conventional levels, if not mentioned otherwise – include, roughly in descending order of absolute coefficient size (in all specifications): previous participation of any household member in a *Hygiene Training* (with an estimated effect size of about ten per cent),[64] and connection either to *Piped Water* only (5 to 6%) or additionally to *Improved Sewerage* (3 to 4%). Exposure to information and communication technology constitutes itself in varying correlation coefficient sizes, depending on which measure is included in the specification. I measure *ICT exposure* using seven alternative variables.[65]

[62] Rather than choosing a single specification as the preferred one, I peg out the range of coefficients of the respective covariates examining a larger array of specifications. I do this in order to triangulate the coefficient size as accurately as possible. When itemising the coefficient ranges in the text, I take into consideration also the specifications presented in the appendix.

[63] Interestingly, when including the variable Sewerage on top of Piped Water, the coefficient of the categorical variable indicating location in the mountain treatment town (reference category is mountain control town) changes its level of statistical significance from the 95-percent level to below 90-percent level.

[64] Marginal effects indicated in parentheses are approximate values. Numbers slightly vary over specifications, mostly only at the third decimal place.

[65] These seven ICT variables are: presence of a *Television Set*, a *Personal Computer*, a *Satellite Dish* antenna (these three do not seem to play a role), a *Radio Receiver*, a *Cell Phone*; connection to a telephone *Landline*, and to the *World Wide Web*.

Table 3: Probability of drinking water treatment, connections compared, selected specifications

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	Probit	Probit	Probit	Probit	Probit	Probit
Piped Water	0.054***	0.033**	0.060***	0.040***	0.061***	0.041***
	(0.012)	(0.015)	(0.013)	(0.015)	(0.014)	(0.015)
Sewerage		0.036**		0.042***		0.042***
		(0.015)		(0.015)		(0.015)
HH Member trained	0.102***	0.096***	0.105***	0.099***	0.106***	0.100***
	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)	(0.021)
Web Access	0.069**	0.066**	0.073**	0.068**	0.073**	0.068**
	(0.033)	(0.033)	(0.033)	(0.034)	(0.033)	(0.033)
Primary Edu	0.044***	0.044***	0.043***	0.044***		
	(0.014)	(0.014)	(0.014)	(0.014)		
Intermediate Edu	0.010	0.008	0.010	0.009		
	(0.023)	(0.022)	(0.022)	(0.022)		
Secondary Edu	0.045**	0.044**	0.050***	0.049***		
	(0.018)	(0.018)	(0.018)	(0.018)		
Tertiary Edu	0.043**	0.042**	0.047**	0.047**		
	(0.021)	(0.020)	(0.021)	(0.021)		
Edu. Durat° HH Head					0.003***	0.003***
					(0.001)	(0.001)
Age HH Head	-0.000	-0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Gender HH Head	-0.044*	-0.042*	-0.047*	-0.045*	-0.042*	-0.040*
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Marital Status HH Head	0.012	0.010	0.011	0.009	0.017	0.015
	(0.021)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)
Disease Awareness	0.006	0.006	0.006	0.006	0.004	0.004
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Asset Index	0.039***	0.038***	0.037***	0.036***	0.038***	0.038***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Dependency Ratio	-0.000	0.004	-0.006	-0.002	-0.010	-0.006
	(0.027)	(0.028)	(0.027)	(0.027)	(0.027)	(0.027)
Water Tank present	-0.021	-0.020	-0.026	-0.026	-0.027	-0.027
	(0.019)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)
Region	0.154***	0.161***				
	(0.016)	(0.017)				
Mountain Treatment			0.035**	0.027	0.036**	0.027
			(0.016)	(0.017)	(0.016)	(0.017)
Coastal Treatment			-0.149***	-0.168***	-0.152***	-0.172***
			(0.024)	(0.025)	(0.024)	(0.025)
Coastal Control			-0.071***	-0.074***	-0.074***	-0.077***
			(0.024)	(0.024)	(0.024)	(0.024)
Sample Size	2459	2459	2459	2459	2459	2459
Pseudo R^2	0.175	0.179	0.186	0.191	0.181	0.186

Average marginal effects displayed. Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

The most influential of these seems to be access to the World Wide Web. Correlation coefficients indicate marginal effect sizes ranging from about three (radio receiver) to seven per cent (web access). The level of *Education* of the household head appears to be an influential determinant as well, with a non-linear course. Compared to the reference category "no education", I estimate completed primary education to increase the probability of treatment of drinking water by four to five per cent. Intermediate education comes with no statistical significance at conventional levels, while the effects of completed secondary and tertiary education each equal that of primary education. I estimate wealth, proxied by the *Asset Index*, to have a marginal effect size of three to four per cent. The *Gender* of the household head being male, in contrast, seems to decrease the probability by four to five per cent.

The estimated coefficients of some explanatory variables – ex ante conceived to have a measurable effect – turn out to lack any statistical significance at conventional levels. *Age* and *Marital Status* of the household head seem to have no influence on the probability of water treatment. The same is true for the respondent knowing about causal relationships of any symptoms of water-related diseases with water quality, sewerage and hygienic conditions (*Disease Awareness*). Likewise, the illness history of household members reflected by *Symptom Occurrence* during the last month before the interview does not appear to be a determinant. The same is true for the *Dependency Ratio*[66] and presence of a *Water Storage Tank*. Interacting *Gender* and *Level of Education* of the household head did not yield any effect.

To test for robustness of results I use different specifications changing variables and estimation procedures. I use varying measures of exposure to ICT, alternative location variables, household composition and education measures, and wealth proxies. Furthermore, I exclude – one at a time – selected variables (*e.g.* using either *Disease Awareness* or *Symptom Occurrence*). The estimations are robust over all specifications for most[67] of the explanatory variables regarding levels of statistical significance, direction of sign and relative coefficient size. Within estimation procedures, also the absolute coefficient sizes remain stable.

Besides 1) *Probit* regressions I also employ 2) *2SPrB* (a two-stage Probit bootstrap, see Appendix 4 for details on this and the following estimators), 3) *RSUR BP* (a recursive version of SUR BP), 4) *IV-Probit*, and 5) *W2SLS* (a specific two-stage procedure proposed by Wooldridge 2010) to control for

[66] The *Dependency Ratio* is statistically significant mainly in specifications without location variables.

[67] There are two exceptions. The level of statistical significance of the variable *Gender of Household Head* drops below the ten per cent convention in a few of the specifications. Presence of a *Water Storage Tank* features coefficients statistically significant at conventional levels only when using the IV-Probit or W2SLS estimator, in each of the cases with a negative sign.

endogeneity. All of these display levels of statistical significance and coefficient signs[68] mostly consistent with all of the basic Probit models for most of the explanatory and control variables. This hints toward the absence of a noteworthy endogeneity bias, but it has to be kept in mind that due to the doubts raised above regarding the fulfilment of the exogeneity condition, these findings are circumstantial evidence at best.

It turns out that some, but not all of the hypotheses can be corroborated. Strong evidence is found that *Hypothesis 3a* (those households in which at least one member participated in any type of hygiene training in the past have a higher likelihood of treating their drinking water) cannot be rejected. Mixed evidence is won regarding *Hypothesis 4a* (those households with exposure to information and communication technology have a higher likelihood of treating their drinking water). While reception of radio broadcasting, use of telecommunication and web access appear to have the expected effects, television seems to be of no importance. This is a surprising subtlety. *Hypothesis 5a* (the more educated the head of the household is, the higher is the likelihood of the household treating its drinking water) receives consistent support.

In contrast to that, neither can *Hypothesis 1a* (connecting a household to piped water decreases its likelihood of treating its drinking water before consumption) nor *Hypothesis 2a* (additional connection of a household to improved sewerage decreases its likelihood of treating its drinking water before consumption, but to a lesser extent than connection to piped water) be sustained. Instead, connected households display a higher propensity to opt for water treatment than the control group. Additional connection to sewerage is positively correlated like connection piped water only, with an effect by comparison about one third smaller. Neither for the household composition nor for the illness history control variables I find results supporting or contradicting the findings of previous literature. The same applies to the role of the age of the household head.

The findings regarding the control variables indicating wealth and female headship are in line with the previous literature, indicating a positive correlation with the likelihood of drinking water treatment.

5.2 Use of cleanser for hand washing

Almost 60 per cent of the households located in the mountainous region report to use cleanser for hand washing, and more than 80 per cent do so in the coastal plain (see Table 4).

[68] Relative coefficient sizes within the IV models do not always reflect those of the Probit model. Some variables seem to gain prediction power on cost of others. The most important covariates in terms of effect size keep their position, though. Absolute coefficient sizes may differ from the marginal effects estimated with the Probit model.

Table 4: Use of cleanser for hand washing, relative frequencies by location and connection group

Connection	Cleanser	Mountain Project	Mountain Control	Total	Coastal Project	Coastal Control	Total
None	Yes	22.97%	51.85%	30.54%	0.00%	75.93%	25.77%
	No	20.93%	48.15%	28.07%	0.00%	24.07%	8.17%
Piped Water	Yes	15.79%	0.00%	11.65%	11.41%	0.00%	7.54%
	No	8.01%	0.00%	5.91%	3.69%	0.00%	2.44%
Sewerage	Yes	23.21%	0.00%	17.12%	72.89%	0.00%	48.15%
	No	9.09%	0.00%	6.71%	12.01%	0.00%	7.93%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
	N	836	297	1133	841	432	1273

Note: no households are connected to piped water or improved sewerage in the control towns. All households in the sample are connected in the coastal project town.

The percentage of compliers (*i.e.*, households that use cleanser) is larger than that of non-compliers in all four towns and three connection status groups (not connected, piped water only, combination of piped water with improved sewerage). It has to be noted that this might partly reflect a positive reporting bias as discussed in the data section. This potential bias is assumed to be non-systematic, *i.e.* not dependent on town or connection status.

Ten major factors seem to determine the probability of a household head in the sample using cleanser when washing hands. I identify nine explanatory variables or groups of variables which are positively correlated with the probability of using cleanser; and one which is negatively correlated.

Keeping the structure of covariates consistent with the overview table presented for the former dependent variable, I oppose the two covariates of main interest in specifications (1) to (6) in Table 5 below. Estimated marginal effects are indicated in parentheses as approximate percentage change of probability of use of cleanser.

Table 13 to Table 16 in Appendix 6 display the estimated marginal effects for complementary specifications. As above, the tables are split by main explanatory variables of interest; and specifications vary media exposure, location, education and household composition variables. Again, for reasons of space limitation alternative specifications and estimators (see Table 15 and Table 16) as well as results of first stages (see Table 29) are presented in the same appendix. The coefficient of determination of the models ranges from eight to eleven per cent.

Table 5: Probability of cleanser use, connections compared, selected specifications

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	Probit	Probit	Probit	Probit	Probit	Probit
Piped Water	0.100***	0.078***	0.124***	0.101***	0.118***	0.096***
	(0.018)	(0.024)	(0.024)	(0.027)	(0.024)	(0.027)
Sewerage		0.037		0.049*		0.048*
		(0.027)		(0.027)		(0.027)
HH Member trained	0.092*	0.087*	0.093*	0.087*	0.090*	0.084*
	(0.050)	(0.050)	(0.050)	(0.050)	(0.049)	(0.049)
Web Access	0.031	0.028	0.032	0.029	0.027	0.023
	(0.084)	(0.084)	(0.083)	(0.083)	(0.083)	(0.083)
Primary Edu	-0.006	-0.007	-0.005	-0.006		
	(0.024)	(0.024)	(0.024)	(0.024)		
Intermediate Edu	0.051	0.048	0.054	0.052		
	(0.039)	(0.039)	(0.039)	(0.039)		
Secondary Edu	0.081***	0.078***	0.084***	0.082***		
	(0.028)	(0.028)	(0.028)	(0.028)		
Tertiary Edu	0.095***	0.093***	0.099***	0.097***		
	(0.032)	(0.032)	(0.031)	(0.031)		
Edu. Durat° HH Head					0.009***	0.009***
					(0.002)	(0.002)
Age HH Head	0.000	0.000	0.000	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gender HH Head	0.002	0.004	-0.001	0.001	-0.016	-0.014
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
Marital Status HH Head	-0.001	-0.002	0.000	-0.002	-0.001	-0.003
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Disease Awareness	0.042**	0.042**	0.042**	0.042**	0.039**	0.039**
	(0.018)	(0.018)	(0.017)	(0.017)	(0.017)	(0.017)
Asset Index	0.044***	0.043***	0.043***	0.042***	0.041***	0.040***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Dependency Ratio	-0.193***	-0.188***	-0.197***	-0.191***	-0.204***	-0.198***
	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
Water Tank present	0.017	0.016	0.012	0.009	0.009	0.007
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Region	-0.183***	-0.177***				
	(0.020)	(0.020)				
Mountain Treatment			0.003	-0.004	0.004	-0.003
			(0.027)	(0.027)	(0.027)	(0.027)
Coastal Treatment			0.154***	0.134***	0.152***	0.132***
			(0.035)	(0.037)	(0.035)	(0.037)
Coastal Control			0.214***	0.210***	0.214***	0.210***
			(0.031)	(0.031)	(0.031)	(0.031)
Sample Size	2466	2466	2466	2466	2466	2466
Pseudo R^2	0.088	0.088	0.089	0.090	0.092	0.093

Average marginal effects displayed. Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

The mentioned variables comprise, approximately in descending order of absolute coefficient size: location in the *Coastal Control* (20 to 21 per cent over reference location *Mountain Control*) or *Project* (18 to 19%) *Town*, connection to the *Piped Water* network (10 to 12%), the *Level of Education* of the household head (in contrast to before though no longer for *Primary Education*, with eight to nine per cent for *Secondary Education* over reference category "*No Education*", and *Tertiary* with nine to ten per cent), participation of a household member in a *Hygiene Training* (9 to 10 %), *Exposure to ICT*[69] (4 to 9%), additional connection to the *Improved Sewerage* network (5%, statistically less significant than *Piped Water*), *Disease Awareness* (4%), *Symptom Occurrence* (3%) and *Assets* as a proxy for wealth (3 to 4%).

Opposing these nine factors is one which is negatively correlated with the probability of using cleaner. An increase of the *Dependency Ratio* by one appears to reduce the probability by up to one fifth (-15 to -20%).

It comes somewhat unexpected that certain factors do not seem to have any influence. These include the presence of a *Water Storage Tank*, or *Primary Education* of the household head – remember I above show both of the latter to be correlated with treatment of drinking water. The same is true for personal characteristics of the household head as *Age*, and *Marital Status*.

It is noteworthy that in contrast to my former dependent variable – probability of treatment of drinking water before consumption – the probability of cleanser use does not appear to be correlated with the *Gender* of the household head.

I check for robustness of the results by estimating different specifications. The estimations are robust over all specifications for almost all of the explanatory variables regarding levels of statistical significance, direction of sign and relative coefficient size. In addition, over different specifications, the absolute coefficient sizes remain stable. Apart from two instrumental variable specifications, in which the location variable coefficient switches sign,[70] there is only one group of variables that requires a closer look. Varying measures of exposure to ICT sheds light on heterogeneous intensities of correlation with cleanser use. While the coefficients of presence of a *Landline, Cell phone,* a *Satellite Dish* or *Radio Receiver* come with statistical significance at the five per cent level or higher, access to the *World Wide Web* does not seem to play any role. Possession of a *Television Set* of a

[69] In contrast to before, presence of a television set is now statistically significant at the ten per cent level, presence of a satellite dish even at the one per cent level. Interestingly, web access is no longer significant at conventional levels. I observe the strongest effect for landline phone connection.

[70] Refer to Kennedy (2005) for a thorough discussion of possible causes of "wrong" signs of estimated coefficients. As the switching of sign occurs only twice, I regard the matter not to be overly concerning.

household is correlated with the probability of use of cleanser stronger and at a higher level of statistical significance when the household has more members[71] than others. One possible explanation is that this finding might reflect the effect of a larger audience (*i.e.*, the television set might be powered during more time of the day, and the household head pick up more information). When the household is connected to *Improved Sewerage* presence of a *Satellite Dish* is significant at the one per cent level, and thus at a higher level than those connected to piped water only (5%).

As before, all of the estimators employed[72] to control for endogeneity display levels of statistical significance and coefficient signs mostly consistent with all of the basic Probit models for most of the explanatory and control variables. As mentioned before in context with the first dependent variable, the Sargan-Hansen test indicates that the instruments do not fulfil the exogeneity condition. Therefore the results of these estimators are of limited informative value.

Deliberating the findings on the probability of cleanser use as a dependent variable, I come to the conclusion that none of my hypotheses – here listed in order of estimated effect sizes of their respective variables – can be rejected. Hypothesis 1b (connecting a household to piped water increases its likelihood of displaying hygiene habits) appears to be correct. So does Hypothesis 5b (the more educated the head of the household is the higher is the likelihood of the household members displaying hygiene habits), for secondary and tertiary education. Also Hypothesis 2b (additional connection of a household to improved sewerage decreases its likelihood of displaying hygiene habits, but to a lesser extent than connection to piped water) seems to reflect reality, including its comparative second component. The same is true for Hypothesis 3b (those households in which at least one member participated in any type of hygiene training in the past have a higher likelihood of displaying hygiene habits). So does Hypothesis 4b (those households with exposure to information and communication technology have a higher likelihood of displaying hygiene habits), while web access and the variable TV and/or PC present constitute the exceptions: they are never statistically significant. Regarding control variables, the picture is mixed. The coefficient of wealth and disease awareness have the expected sign, while larger households in contrast to my expectation seem to have a lower probability to use cleanser than smaller ones. Neither recent symptom occurrence, nor age or gender of the household head seems to play a role.

[71] This is respectively reflected by the following household composition variables: number of children, number of children and elderly, and total number of household members.

[72] The same four instrumental variable estimators are used as for the dependent variable *Probability of drinking water treatment*: *2SPrB*, *RSUR BP*, *IV-Probit*, and *W2SLS.*

As an additional approach to establish robustness of the findings; and to increase confidence that the probability of cleanser use is indeed a good proxy for hygiene habits, I regress three alternative continuous dependent variables on the covariates: *Spending on Soap*, *Spending on Cleanser*, and *Frequency of Latrine-cleaning.* The findings are largely[73] consistent with those of the present section. I describe this in detail in Appendix 6. The mentionable exceptions are that *Improved Sewerage* is not statistically significant for the dependent variable *Spending on Soap*; and both connection types are not statistically significant for the dependent variable *Spending on Cleanser*. In contrast to *Piped Water*, *Improved Sewerage* is not statistically significant for the dependent *Variable Latrine Cleaning Frequency*. The variable *Hygiene Training* is not statistically significant at conventional levels for the dependent variable *Spending on Soap*, but for all four other dependent variables, including *Spending on Cleanser*.

6. Conclusion

The purpose of this paper is to identify the determinants of water treatment and hygiene habits in households in Yemenite provincial towns. I do this by empirically analysing micro-level data collected during a large-scale household survey; and summarize the main findings as follows.

The probability of *Drinking Water Treatment* is positively correlated with location of the household in the mountainous region, hygiene training experience, connection to piped water, additional connection to improved sewerage, ICT exposure, and wealth, completed primary, secondary and tertiary education. The gender of the household head being male, and the household being located in the coastal plain – especially in the project town, is negatively correlated.

I use four different measures to map *Hygiene Habits*, *i.e.* the reported practices and resource allocation on household level linked to personal hygiene and health-relevant aspects of housekeeping. The measures are not chosen as stand-alone signifiers of hygiene habits, but are intended to allow composing a more comprehensive and detailed picture of determinants of habitual hygiene behaviour. The analysis reveals that connection to piped water is a more important driver than that to improved sewerage. Participation in hygiene training is among the most strongly correlated covariates. Wealth and recent affection foster hygiene habits. The same I can say about ICT, while I have to challenge the idea that the underlying mechanism is entirely rooted in better access to information. In the following, the four outcome variables are crisply covered one by one.

[73] The variable hygiene training is not statistically significant at conventional levels for the dependent variable spending on soap, but for all four other dependent variables, including spending on cleanser.

First, the probability of *Use of Cleanser* for hand washing appears to be fostered by location in the coastal region, connection to piped water, level of education, hygiene training experience, connection to improved sewerage, ICT exposure, disease awareness, recent illness history and wealth. The dependency ratio is negatively correlated. Second, *Spending on Soap* is positively correlated with ICT exposure, location in the mountainous region, and connection to piped water, use of a water storage tank, wealth and recent illness history. Location in one of the coastal towns and age of the household head are negatively correlated with spending. Third, *Spending on Cleanser* (including laundry detergent) seems to be increased by ICT exposure, hygiene training experience, location in the mountainous region, disease awareness, recent illness history, connection to piped water (only in few of the specifications) and wealth. The dependency ratio, location in one of the coastal towns and age of the household head are negatively correlated. Fourth, the *Frequency of Latrine-Cleaning* appears to be driven by mean education duration of adults, training, the dependency ratio, connection to piped water, wealth, location in the coastal project town, presence of a water storage tank, and education. On the other hand, location in the coastal control town or in the mountainous region is negatively correlated with the frequency.

It requires some thought how the findings of the paper at hand can be squared with those of Klasen *et al.* (2012); and consecutive work by Lechtenfeld (2012). The former carve out that piped water deteriorates health (in the mountainous area) or has so significant effect (in the coastal area); and improved sewerage has no significant (mountainous) to marginally improving (coastal) health effects. We find that incidence of e.coli is high in treatment households, with more than half of it stemming from leaking pipes and rationing; and the remainder from household behaviour. The findings of Lechtenfeld (2012) are in line with this, stating that access to piped water increases the "risk of child diarrhea by 4.6 percentage points"; with the majority of the pollution stemming from "broken pipes and interruptions of water supply", exacerbated by "unhygienic water storage and handling at household level". In This paper I now find that connection to piped water and also to improved sewerage are positively correlated with the probability of drinking water treatment and reporting of following hygiene habits. While this seems to contradict the detrimental to at best feeble health impact of connection to piped water found by Klasen et al. (2012); and by Lechtenfeld (2012), three reconciling explanations are conceivable: First, although connected households seem to display a higher probability of treating their drinking water and of displaying hygiene habits, the beneficial effect of this behaviour on health might be overcompensated by the detrimental effect of the connection on health. Second, water treatment at the household level might – although implemented – be ineffective for some reason, possibly faulty practices. It seems less likely to me that poor conduct of hygiene practices might be an issue here, as water treatment is more technical and complicated than body and living quarter's hygiene. Third, the behaviour might be an

adjustment reaction of connected households realizing that the connection might increase the risk of contamination of the drinking water at point-of-use. All three explanations underline that pollution caused on the supply side is a major concern.

The study at hand faces two noteworthy limitations. First, all data used in the analysis is self-reported information about behavioural aspects regarding various aspects of hygiene. Yemenite female enumerators talking face to face for protracted time to women in the survey households conducted the interviews. The potential concern that reporting could be biased in favour of behaviour deemed by the respondents to meet the expectations of the enumerator cannot entirely be dismissed. Craving for admiration might palliate the picture. Anyway, a good case can be made why not to have to worry too much about this potential bias: there is no reason to believe that it should be systematic in any way, i.e. affecting households in different ways depending on their location, social strata et cetera. Neither did the interviewers report any peculiarities in this regard, nor did searching for patterns in the data reveal anything hinting towards such systematic differences. Second, limited subsample sizes restrict the explanatory power of the study regarding resource allocation. For instance, data on expenditure on cleanser is available for no more than 118 households[74] when analysing those located in the coastal plain and connected to piped water, but not improved sewerage. Nothing can be done to remedy this difficulty.

As presented in the introduction, Yemen is similar to many MENA countries – and beyond – in terms of topography, hydrography, demography, public health and socio-economic development faced. Admittedly, some of these circumstances are posing challenges of a more complex and urgent nature in Yemen than elsewhere. However, conceptually comparable to the effect of a magnifying glass, due to these intensifying circumstances – increasing the visibility of need for action – the country might so much the better serve as a case study from which lessons can be learned for other countries and societies. Mechanisms at work are presumed to become evident more articulate in the tightened Yemenite setting. External validity can be insinuated inasmuch as those determinants identified in this study to drive behaviour will probably be at work in other locations as well, albeit in varying magnitudes of effect.

Several policy implications suggest themselves. I identify a couple of determinants of hygiene-related behaviour that planners and designers of infrastructure projects in the water and sanitation section can take into account. Furthermore, some of these determinants can be influenced in the medium

[74] This is the lowest subsample size in this study. Data from the section of the questionnaire used to survey data on consumption expenditure is less exhaustive than that from other sections. Part of this is rooted in a certain reluctance demonstrated by many Yemenite households to disclose this particular kind of information.

term in order to achieve more favourable health outcomes of such projects than up to now. Especially flanking and complementing construction of piped water and sewerage from the outset with hygiene training offers and health awareness campaigns networks – either based on multipliers or with blanket coverage – shows promise.

Future research should address two main fields of interest. First, to my knowledge, nobody conducted Randomized Controlled Trial (RCT) or quasi-experimental studies yet in order to establish the optimal level of intensity, resource allocation and mode of conduct regarding hygiene trainings and health awareness campaigns. This knowledge gap should be filled in order to improve integrated development programme planning and cost-benefit ratios. Second, the role of ICT in shaping and influencing perceptions, beliefs and habits regarding water handling practices and personal hygiene should be inspected more closely. In particular, the shifting relevance of different media and devices deserves attention. Gaugeable data and information should be attained on the role of the World Wide Web, social media and modern mobile telecommunication devices. Apart from that, the questions analysed in the study at hand should be researched in other geographic regions, with samples of sufficient size to ensure validity of findings for all sub-groups of the surveyed population.

Bibliography

Adkins, Lee C. (2009): An Instrumental Variables Probit Estimator using gretl. In Ignacio Diaz-Emparanza, Petr Mariel, Maria Victoria Esteban (Eds.): Econometrics with gretl. Proceedings of the gretl conference 2009. Bilbao.

Ai, Chunrong; Norton, Edward C. (2003): Interaction terms in logit and probit models. In *Economics Letters* 80 (1), pp. 123–129.

Aiello, Allison E.; Coulborn, Rebecca M.; Perez, Vanessa; Larson, Elaine L. (2008): Effect of hand hygiene on infectious disease risk in the community setting: a meta-analysis. In *American Journal of Public Health* 98 (8), pp. 1372–1381.

Angrist, Joshua D. (2001): Estimation of limited dependent variable models with dummy endogenous regressors. In *Journal of Business & Economic Statistics* 19 (1), pp. 2–16.

Angrist, Joshua D.; Pischke, Jörn-Steffen (2008): Mostly harmless econometrics: An empiricist's companion. Princeton, NJ: Princeton University Press.

Angrist, Joshua; Krueger, Alan B. (2001): Instrumental variables and the search for identification: From supply and demand to natural experiments. In *Journal of Economic Perspectives* 15 (4), pp. 69–85.

Aunger, Robert; Schmidt, Wolf-Peter; Ranpura, Ashish; Coombes, Yolande; Maina, Peninnah Mukiri; Matiko, Carol Nkatha; Curtis, Valerie (2010): Three kinds of psychological determinants for hand-washing behaviour in Kenya. In *Social Science and Medicine* 70 (3), pp. 383–391.

Badowski, Natalie; Castro, Cynthia M.; Montgomery, Maggie; Pickering, Amy J.; Mamuya, Simon; Davis, Jennifer (2011): Understanding household behavioral risk factors for diarrheal disease in Dar es Salaam: A photovoice community assessment. In *Journal of Environmental and Public Health* 2011.

Barry, Michele; Hughes, James M. (2008): Talking dirty – the politics of clean water and sanitation. In *New England Journal of Medicine* 359 (8), pp. 784–787.

Bartram, Jamie; Lewis, Kristen; Lenton, Roberto; Wright, Albert (2005): Focusing on improved water and sanitation for health. In *The Lancet* 365 (9461), pp. 810–812.

Begum, Sharifa; Ahmed, Mansur; Sen, Binayak (2011): Do Water and Sanitation Interventions Reduce Childhood Diarrhoea? New Evidence from Bangladesh. In *Bangladesh Development Studies* 34 (3), pp. 1–30.

Berkman, Douglas S.; Lescano, Andres G.; Gilman, Robert H.; Lopez, Sonia L.; Black, Maureen M. (2002): Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. In *The Lancet* 359 (9306), pp. 564–571.

Bernheim, B. Douglas; Garrett, Daniel M.; Maki, Dean M. (2001): Education and saving:: The long-term effects of high school financial curriculum mandates. In *Journal of Public Economics* 80 (3), pp. 435–465.

Brett, Eduard A. (2003): Participation and accountability in development management. In *Journal of Development Studies* 40 (2), pp. 1–29.

Brikké, François; Bredero, Maarten (2003): Linking technology choice with operation and maintenance in the context of community water supply and sanitation. Reference document for planners and project staff. World Health Organization and IRC International Water and Sanitation Centre. Geneva.

Burton, Maxine; Cobb, Emma; Donachie, Peter; Judah, Gaby; Curtis, Val; Schmidt, Wolf-Peter (2011): The effect of handwashing with water or soap on bacterial contamination of hands. In *International Journal of Environmental Research and Public Health* 8 (1), pp. 97–104.

Chase, Claire; Do, Quy Toan (2012): Handwashing behavior change at scale: evidence from a randomized evaluation in Vietnam. World Bank (Policy Research Working Paper, 6207).

Chiburis, Richard; Das, Jishnu; Lokshin, Michael (2011): A Practical Comparison of the Bivariate Probit and Linear IV Estimators. World Bank (Policy Research Working Paper, 5601).

Curtis, Val; Cairncross, Sandy (2003): Effect of washing hands with soap on diarrhoea risk in the community: a systematic review. In *The Lancet Infectious Diseases* 3 (5), pp. 275–281.

Curtis, Val; Schmidt, Wolf; Luby, Stephen; Florez, Rocio; Touré, Ousmane; Biran, Adam (2011): Hygiene: new hopes, new horizons. In *The Lancet Infectious Diseases* 11 (4), pp. 312–321.

Davis, Jennifer; Pickering, Amy J.; Rogers, Kirsten; Mamuya, Simon; Boehm, Alexandria B. (2011): The effects of informational interventions on household water management, hygiene behaviors, stored drinking water quality, and hand contamination in peri-urban Tanzania. In *American Journal of Tropical Medicine and Hygiene* 84 (2), p. 184.

Devine, Jacqueline (2010): Beyond tippy-taps: The role of enabling products in scaling up and sustaining handwashing. In *Waterlines* 29 (4), pp. 304–314.

Dong, Yingying; Lewbel, Arthur (2012): A simple estimator for binary choice models with endogenous regressors. Boston College (Boston College Working Papers in Economics, 807). Available online at fmwww.bc.edu/EC-P/wp807.pdf.

Dreher, Axel; Klasen, Stephan; Raymond, James; Werker, Eric (2010): The costs of favoritism: Is politically-driven aid less effective? Center for European, Governance and Economic Development Research, University of Göttingen (CeGE Discussion Papers, 97).

Dupas, Pascaline (2011): Health behavior in developing countries. In *Annual Review of Economics* 3 (1), pp. 425–449.

Ebrahim, Alnoor (2003): Accountability in practice: mechanisms for NGOs. In *World Development* 31 (5), pp. 813–829.

Ejemot-Nwadiaro, Regina I.; Ehiri, John E.; Meremikwu, Martin M.; Critchley, Julia A. (2012): Hand washing for preventing diarrhoea (Review). In *The Cochrane Library* (2), pp. 1–39.

Feil, E. J. (2012): The emergence and spread of dysentery. In *Nature Genetics* 44 (9), pp. 964–965.

Flemming, Hans-Curt (2009): Marine and Industrial Biofouling. Why Microorganisms live in biofilms and the problem of biofouling. In *Springer Series on Biofilms* (4), pp. 3–12.

Freedman, David A.; Sekhon, Jasjeet S. (2010): Endogeneity in probit response models. In *Political Analysis* 18 (2), pp. 138–150.

Galiani, Sebastian; Gertler, Paul; Orsola-Vidal, Alexandra (2012): Promoting handwashing behavior in Peru: the effect of large-scale mass-media and community level interventions. Impact Evaluation Series No. 74. World Bank (Policy Research Working Paper, 6257).

Greene, William H. (2012): Econometric Analysis. 7th ed. Upper Saddle River, NJ: Prentice hall.

Haidera, Mansour; Alhakimi, Saif Ali; Noaman, Abdulla; Al Kebsi, Alkhatib; Noaman, Anwar; Fencl, Amanda; Dougherty, Bill; Swartz, Chris (2011): Water scarcity and climate change adaptation for Yemen's vulnerable communities. In *Local Environment* 16 (5), pp. 473–488.

Hardin, James W. (2012): Bivariate probit models. In *Stata Technical Bulletin* 6 (33), pp. 15–20.

Heckman, James J. (1978): Dummy endogenous variables in a simultaneous equation system. In *Econometrica* 46 (4), pp. 931–959.

Heckman, James J. (1979): Sample selection bias as a specification error. In *Econometrica* 47 (1), pp. 153–161.

Holm, Anders; Arendt, Jacob N. (2012): Evaluating the performance of simple estimators for probit models with two dummy endogenous regressors. In *Journal of Statistical Computation and Simulation* (ahead-of-print), pp. 1–23.

Jensen, Peter Kjaer; Ensink, Jeroen H. J.; Jayasinghe, Gayathri; van der Hoek, Wim; Cairncross, Sandy; Dalsgaard, Anders (2002): Domestic transmission routes of pathogens: the problem of in-

house contamination of drinking water during storage in developing countries. In *Tropical Medicine and International Health* 7 (7), pp. 604–609.

Kennedy, Peter E. (2005): Oh no! I got the wrong sign! What should I do? In *Journal of Economic Education* 36 (1), pp. 77–92.

Klasen, Stephan; Lechtenfeld, Tobias; Meier, Kristina; Rieckmann, Johannes P. (2011): Water Supply and Sanitation in Provincial Towns in Yemen. Impact Evaluation Report. Courant Research Centre 'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis', University of Göttingen (Discussion Paper, 102).

Klasen, Stephan; Lechtenfeld, Tobias; Meier, Kristina; Rieckmann, Johannes P. (2012): Benefits trickling away: the health impact of extending access to piped water and sanitation in urban Yemen. In *Journal of Development Effectiveness* 4 (4), pp. 537–565.

Kosek, Margaret; Bern, Caryn; Guerrant, Richard L. (2003): The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. In *Bulletin of the World Health Organization* 81 (3), pp. 197–204.

Kremer, Michael; Miguel, Edward (2007): The illusion of sustainability. In *Quarterly Journal of Economics* 122 (3), pp. 1007–1065.

Lange, W. J. de; Mahumani, B. K.; Steyn, M.; Oelofse, S. H. H. (2012): Monetary valuation of salinity impacts and microbial pollution in the Olifants Water Management Area, South Africa. In *Water SA* 38 (2), pp. 241–248.

Lechtenfeld, Tobias (2012): Vulnerability to poverty: determinants and policy issues. 1st ed. Göttingen: Optimus Wissenschafts-Verlag.

Luby, Stephen P.; Agboatwalla, Mubina; Painter, John; Altaf, Arshad; Billhimer, Ward; Keswick, Bruce; Hoekstra, Robert M. (2006): Combining drinking water treatment and hand washing for diarrhoea prevention, a cluster randomised controlled trial. In *Tropical Medicine and International Health* 11 (4), pp. 479–489.

Luoto, Jill E.; Levine, David; Albert, Jeff (2011): Information and persuasion: achieving safe water behaviors in Kenya. RAND. Santa Monica, CA (Labor and Population Working Paper, WR-885).

Maddala, Gangadharrao S. (1983): Limited-dependent and qualitative variables in econometrics. Cambridge, MA: Cambridge University Press.

Murray, Christopher J. (1994): Quantifying the burden of disease: the technical basis for disability-adjusted life years. In *Bulletin of the World Health Organization* 72 (3), p. 429.

Newey, Whitney K. (1987): Efficient estimation of limited dependent variable models with endogenous explanatory variables. In *Journal of Econometrics* 36, pp. 231–250.

Niyogi, Swapan Kumar (2005): Shigellosis. In *Journal of Microbiology* 43 (2), pp. 133–143.

Parker, Amy A.; Stephenson, R.; Riley, P. L.; Ombeki, S.; Komolleh, C.; Sibley, L.; Quick, Robert (2006): Sustained high levels of stored drinking water treatment and retention of hand-washing knowledge in rural Kenyan households following a clinic-based intervention. In *Epidemiology and Infection* 134 (5), pp. 1029–1036.

Park, Hun Myoung (2010): Regression Models for Binary Dependent Variables Using Stata, SAS, R, LIMDEP, and SPSS. Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University. Bloomington, IN.

Patel, Ronak B.; Stoklosa, Hanni; Shitole, Shrutika; Shitole, Tejal; Sawant, Kiran; Nanarkar, Mahesh; Subbaraman, Ramnath; Ridpath, Alison; Patil-Deshmuk, Anita (2013): The high cost of diarrhoeal illness for urban slum households–a cost-recovery approach: a cohort study. In *The World Bank Research Observer* 3 (4).

Pengpid, Supa; Peltzer, Karl (2012): Hygiene behaviour and health attitudes in African countries. In *Current Opinion in Psychiatry* 25 (2), pp. 149–154.

Prüss-Üstün, Annette; Corvalán, Carlos (2006): Preventing disease through healthy environments. Towards an estimate of the environmental burden of disease. Geneva.

Rosen, Sydney; Vincent, Jeffrey R. (1999): Household water resources and rural productivity in Sub-Saharan Africa: A review of the evidence. Harvard Institute for International Development, Harvard University. Cambridge, MA.

Santosham, Mathuram; Chandran, Aruna; Fitzwater, Sean; Fischer-Walker, Christa; Baqui, Abdullah H.; Black, Robert (2010): Progress and barriers for the control of diarrhoeal disease. In *The Lancet* 376 (9734), pp. 63–67.

Saretsky, Gary (1972): The OEO PC experiment and the John Henry effect. In *The Phi Delta Kappan* 53 (9), pp. 579–581.

Sattar, Abdul; Ahmad, Eatzaz; Pant, Krishna Prasad (2007): Willingness to Pay for the Quality of Drinking Water [with Comments]. In *Pakistan Development Review* 46 (4), pp. 767–777.

Schmidt, Wolf-Peter; Aunger, Robert; Coombes, Yolande; Maina, Peninnah Mukiri; Matiko, Carol Nkatha; Biran, Adam; Curtis, Val (2009): Determinants of handwashing practices in Kenya: the role of media exposure, poverty and infrastructure. In *Tropical Medicine and International Health* 14 (12), pp. 1534–1541.

Stock, James H.; Yogo, Motohiro (2005): Testing for Weak Instruments in Linear IV Regression. In Donald W. K. Andrews, James H. Stock (Eds.): Identification and Inference for Econometric Models. Essays in Honor of Thomas Rothenberg. Cambridge, UK: Cambridge University Press, pp. 80–108.

Sugita, Eri Woods (2004): Domestic water use, hygiene behavior, and children's diarrhea in rural Uganda. Doctoral dissertation. University of Florida.

Taher, T.; Bruns, B.; Bamaga, O.; Al-Weshali, A.; van Steenbergen, F. (2012): Local groundwater governance in Yemen: building on traditions and enabling communities to craft new rules. In *Hydrogeology Journal* 20 (6), pp. 1177–1188.

Tan Jr., Carlos Antonio R.; Capuno, Joseph J. (2012): Household decisions and child health: Estimating the links between water treatment and the incidence of diarrhea using non-recursive two-equation causal models. University of the Philippines. Philippines (UP School of Economics Discussion Papers, 2012-02).

Terza, Joseph V. (1986): Determinants of household electricity demand: A two-stage probit approach. In *Southern Economic Journal*, pp. 1131–1139.

Totouom, Armand Luc Fotue; Sikod, Fondo; Abba, Ibrahim (2012): Household choice of purifying drinking water in Cameroon. In *Environmental Management and Sustainable Development* 1 (2), pp. 101–115.

Trevett, Andrew; Carter, Richard; Tyrrel, Sean (2005): The importance of domestic water quality management in the context of faecal-oral disease transmission. In *Journal of Water and Health* 3 (3), pp. 259–270.

Turkington, Darrell A.; Bowden, R. J. (1988): Identification information and instruments in linear econometric models with rational expectations. In *Journal of Econometrics* 38 (3), pp. 361–373.

United Nations (2012): World Urbanization Prospects: The 2011 Revision, CD-ROM Edition. File 6: Average Annual Rate of Change of the Urban Population by Major Area, Region and Country, 1950-2050 (per cent). Department of Economic and Social Affairs, Population Division. New York.

United Nations General Assembly (2010): Resolution 64/292: The human right to water and sanitation. A/RES/64/292.

Waddington, Hugh; Snilstveit, Birte (2009): Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea. In *Journal of Development Effectiveness* 1 (3), pp. 295–335.

Walker, Christa L. F.; Rudan, Igor; Liu, Li; Nair, Harish; Theodoratou, Evropi; Bhutta, Zulfiqar A.; O'Brien, Katherine L.; Campbell, Harry; Black, Robert E. (2013): Global burden of childhood pneumonia and diarrhoea. In *The Lancet* 381 (9875), pp. 1405–1416.

Wickström, Gustav; Bendix, Tom (2000): The" Hawthorne effect"—what did the original Hawthorne studies actually show? In *Scandinavian Journal of Work, Environment and Health* 26 (4), pp. 363–367.

Wooldridge, Jeffrey M. (2010): Econometric Analysis of Cross-Sectional and Panel Data. 2nd edition. Cambridge, MA: The MIT Press.

World Bank (2005): GINI index. The World Bank. Available online at http://data.worldbank.org/indicator/SI.POV.GINI, last accessed July 30, 2013.

World Bank (2008): What Works in Water Supply and Sanitation? Lessons from Impact Evaluations. A Summary of Findings. IEG Independent Evaluation Group, The World Bank.

World Bank (2010): Water and Development. An Evaluation of World Bank Support, 1997-2007. Independent Evaluation Group (World Bank, IFC, MIGA). Washington, DC (IEG Study Series, 1 & 2).

World Bank (2013a): Annual population growth data. The World Bank. Available online at http://data.worldbank.org/indicator/SP.POP.GROW/countries, last accessed July 30, 2013.

World Bank (2013b): World Development Report 2013: Jobs. The World Bank. Washington, DC. Available online at http://siteresources.worldbank.org/EXTNWDR2013/Resources/8258024-1320950747192/8260293-1322665883147/WDR_2013_Report.pdf.

World Health Organization (2008): Guidelines for Drinking-water Quality: third Edition incorporating the first and second Addenda. Volume 1, Recommendations. Geneva.

Zeug, G.; Eckert, S. (2010): Population Growth and Its Expression in Spatial Built-up Patterns: The Sana'a, Yemen Case Study. In *Remote Sensing* 2 (4), pp. 1014–1034.

Zomer, Tizza P.; Erasmus, Vicki; van Empelen, Pepijn; Looman, Caspar; van Beeck, Ed F. (2013): Sociocognitive determinants of observed and self-reported compliance to hand hygiene guidelines in child day care centers. In *American Journal of Infection Control* 41 (6), pp. 497–502.

Zwane, A.P; Kremer, M. (2007): What Works in Fighting Diarrheal Diseases in Developing Countries? A Critical Review. In *The World Bank Research Observer* 22 (1), pp. 1–24.

Appendix 1: Technical notes

Methods of water purification differ in their capability to withhold or neutralize different contaminations (viruses, bacteria, vermin, chemicals, excess or harmful minerals like arsenic or fluorides, and suspended particles), their speed of use and suitability for different volumes, and their effectiveness to improve colour, odour and taste of the water. These methods include domestic chemical disinfection using chlorine or iodine, boiling, filtration, pre-settlement before use, storage and sedimentation, a coagulation-flocculation-settlement-procedure, aeration, exposure to sunlight, and desalination (see Brikké, Bredero 2003Brikké and Bredero, 2003). Below, I summarize the most important details regarding these methods.

While *chlorination* is easy and cost-efficient, its drawback is its unsavoury taste. The chlorine can be added in form of powder, tablets or liquid. Chlorine compounds immediately start to react with pathogen in the water. If enough chlorine was added, some free residual chlorine will remain detectable after all pathogens have been destroyed. A concentration of 0.5 milligram per litre is recommended (World Health Organization, 2008) in order to secure continued disinfection in case new pathogens enter, while higher concentrations will shift the trade-off between disinfection and taste to the detriment of the latter. Initial water quality – including turbidity – commands the amount of chlorine needed. *Iodine* has a similar effect, while the taste is perceived by many users to be less offensive, especially when adding vitamin C to the water following the contact time. Nonetheless, water treated with iodine is not recommended for consumption for pregnant women and persons allergic to iodine.

Boiling reliably kills off microorganisms causing diarrheal diseases, even at high altitudes with low air pressure and a lowered boiling point of the water. As the effect is reached already below the boiling point, even short ebullition is sufficient.

Filtration works through particle removal or adsorption. Several methods are available, differing in level of sophistication, cost and speed. These include slow and rapid sand filtration, straining through fine cloth, ceramic filters, silver-ion filters, membrane and reverse osmosis processes, combined technology filtration, and adsorption using granular charcoal or activated carbon filters.

Simple *presettlement* of the drinking water for more than two days eliminates besides some bacteria also *Carcaria* snail larvae, for example. Longer *storage and sedimentation* is a method rather used at communal than at domestic level due to the rather time-consuming character of the procedure, which makes it inconvenient for day-to-day application. Solid contaminants sink to the ground of the storage container and can then be separated from the drinking water. When employing the *coagulation-flocculation-settlement-procedure*, a liquid coagulant is intermixed with the water in

order to bind particles. These then cohere to larger flocks and collect at the ground or can be removed otherwise. Aluminium sulphate is an example for a suitable coagulant. *Aeration* oxidizes iron and manganese.

Exposure to sunlight in a clear plastic – not glass – bottle to direct ultraviolet rays will kill the majority of microbes in the water within a couple of hours. The method is sometimes referred to as solar water disinfection (SODIS).

Desalination alleviates disinfection due to "low total organic carbon and particle content, low microbial loads and minimal oxidant demand" (Curtis, Cairncross 2003), but should be combined with chlorination.

Appendix 2: Description of covariates

Table A.1: Description of covariates

Explanatory variables	Type	Description
Main explanatory variables		
Piped Water	Binary	Indicator if the household is connected to the piped water network (1=yes, 0=no)
Sewerage	Binary	Indicator if the household is connected to the improved sewerage network, conditional on connection to piped water (1=yes, 0=no)
Household head characteristics		
Age of HH Head	Continuous	Age of the household head in years
Gender of HH Head	Binary	Indicator of gender of household head (male=1, female=0)
Marital Status HH Head	Binary	Indicator if household head is currently married (1=yes, 0=no)
Edu. of HH Head: None	Categorical	Indicator of level of education completed by household head: none
Edu. of HH Head: Primary	Categorical	Indicator of level of education completed by household head: primary
Edu. of HH Head: Middle	Categorical	Indicator of level of education completed by household head: middle/ preparatory
Edu. of HH Head: Secondary	Categorical	Indicator of level of education completed by household head: secondary
Edu. of HH Head: Tertiary	Categorical	Indicator of level of education completed by household head: tertiary/ university
Edu. Duration HH Head	Continuous	Number of years of education the household head enjoyed
Literacy of HH Head	Binary	Indicator if the household head is capable of reading (1=yes, 0=no)
Household characteristics		
Dependency Ratio	Continuous	Ratio of number of household members outside of working age bracket 15-60 and those inside, ranges from 0 to 1
# of HH Members	Continuous	Number of all household members
# of Children	Continuous	Number of children (up to age of twelve years)
# Children & Elderly	Continuous	Number of children (up to age of twelve years or older than 70)
Mean Edu. Duration Adults	Continuous	Average number of years of education an adult in a given household has enjoyed
Disease Awareness	Binary	Indicator taking the value 1 if the respondent correctly reports 5 symptoms to possibly be related with water handling and hygiene
Symptom Occurrence	Binary	Indicator if any household members experienced any symptoms during the past four weeks before the interview (1=yes, 0=no)
HH Member trained	Binary	Indicator if any household member previously participated in any type of hygiene training (1=yes, 0=no)
Asset Index	Continuous	Aggregate index of wealth based on reported housing characteristics and selected assets, by region, polychoric PCA
Water Tank present	Binary	Indicator if the household uses any water tanks to store water (1=yes, 0=no)
Soap bought	Binary	Indicator if there is at least one piece of soap bought during a typical month (1=yes, 0=no)

Location effects		
Region	Binary	Indicator of location in mountainous (1) or coastal (0) region
Mountain Treatment	Binary	Indicator of location in mountainous project town (1=yes, 0=no)
Mountain Control	Binary	Indicator of location in mountainous control town (1=yes, 0=no)
Coastal Treatment	Binary	Indicator of location in coastal project town (1=yes, 0=no)
Coastal Control	Binary	Indicator of location in coastal control town (1=yes, 0=no)
Exposure to ICT and mass media		
TV present	Binary	Indicator if the household owns a television set (1=yes, 0=no)
TV and/or PC present	Binary	Indicator if the household owns a television set and/ or a personal computer (1=yes, 0=no)
Sat. Dish present	Binary	Indicator if the household owns a satellite dish (1=yes, 0=no)
Radio present	Binary	Indicator if the household owns a radio broadcasting receiver (1=yes, 0=no)
GSM present	Binary	Indicator if at least one cell phone is owned by one of the household members (1=yes, 0=no)
Landline Connection	Binary	Indicator if the household owns a landline phone (1=yes, 0=no)
Web Access	Binary	Indicator if the household has access to the World Wide Web (1=yes, 0=no)
Housing characteristics		
IV Distance to Centre	Continuous	Instrumental variable, indicating distance of house from the centre of the town in meters (based on GPS data)
IV House on Rock	Binary	Instrumental variable, indicating if the house rests upon shapeable (0) or rocky subgrade (1)
IV Age of House	Continuous	Instrumental variable, indicating reported age of the house (ranging from new to 820 years)
Building material 1	Categorical	Indicator that the outer walls of the house consist of straw, sticks and/ or mud
Building material 2	Categorical	Indicator that the outer walls of the house consist of straw, sticks and/ or mud wood, mud and plastering
Building material 3	Categorical	Indicator that the outer walls of the house consist of natural stone
Building material 4	Categorical	Indicator that the outer walls of the house consist of sun-dried bricks
Building material 5	Categorical	Indicator that the outer walls of the house consist of burnt bricks
Building material 6	Categorical	Indicator that the outer walls of the house consist of stone and cement
Building material 7	Categorical	Indicator that the outer walls of the house consist of cement

Appendix 3: Descriptive statistics

Table A.2: Means and differences of outcomes by connection status

Outcome	Mean, not connected (control)	Min	Max	Mean, piped water	Min	Max	Diff. to control	Mean, improved sewerage	Min	Max	Diff. to control
1) Treatment of drinking water	0.069	0	1	0.120	0	1	0.050***	0.107	0	1	0.037***
2) Use of cleanser for hand washing	0.614	0	1	0.695	0	1	0.083***	0.820	0	1	0.208***
3) Spending for soap	64.85	0	500	84.46	0	428.89	19.61***	82.97	0	500	18.12***
4) Spending for cleanser	144.78	0	760	149.42	0	562.50	4.64	162.88	0	833.33	18.10***
5) Frequency of latrine-cleaning	7.57	0	28	7.93	1	21	0.37*	8.38	0	35	0.82***

Table A.3: Mean spending (per capita, in YER) on cleanser by location and connection status

Connection	Mountain Project (YER)	N	SD	Mountain Control (YER)	N	SD	Coastal Project (YER)	N	SD	Coastal Control (YER)	N	SD
None	153.87	265	111.93	160.01	196	123.26	-	-	-	131.85	425	116.45
Piped Water	159.38	125	105.01	-	-	-	138.91	118	115.96	-	-	-
Sewerage	173.21	180	119.26	-	-	-	160.26	660	132.60	-	-	-

N over all connection groups and locations: 1969

Note: no households are connected to piped water or improved sewerage in the control towns. All households in the sample are connected in the coastal project town. Values capped at 500 YER per capita for each component (spending on soap, spending on laundry detergent) to exclude industrial use.

Appendix 4: Methodology of employed instrumental variable approaches

I briefly summarize the findings on relevance and validity of the instrumental variables considered for use in This paper. For the dependent variable *Probability of Drinking Water Treatment*, the instruments *IV Age of House* and *IV Rocky Ground* used in the Wooldridge procedure (which is explained below) appear valid with F-statistics of the predicted values well above the critical value[75] for both (separately) instrumented covariates. Exogeneity is more of a concern, as the Sargan-Hansen test for overidentification suggests that the null hypothesis – which says that the instrument or set of instruments is valid – can be rejected on grounds of statistical significance[76] at the conventional level of five (when instrumenting *Connection to Piped Water*) respectively ten (*Connection to Improved Sewerage*) per cent.[77] For the dependent variable *Use of Cleanser for Hand Washing* the same two instruments appear valid, while the Sargan-Hansen test again does not provide support for my assumption of exogeneity.[78]

For *Spending on Soap* the picture looks similarly unsatisfactory. The two instruments receive less support regarding their validity now, while the null of exogeneity of the instruments can still be rejected at the ten per cent level.[79] When introducing the third instrument *IV Distance* on top of the other two, validity is compellingly supported again; but lack of support of the assumption of exogeneity by the Sargan-Hansen test remains an issue.[80] For the more comprehensive *Spending on*

---

[75] For the instrumented covariate *Connection to Piped Water* the Cragg-Donald Wald F-statistic is 36.90 compared to a Stock-Yogo critical reference value of 16.38 for one instrumented endogenous covariate and one instrument, and a desired size of a Wald test of 0.10 at a significance level of five per cent. For the covariate *Connection to Improved Sewerage* the F-statistic reads 30.54. Note that the two instruments are reduced to one in the Wooldridge procedure, namely the predicted value of the instrumented connection indicator.

[76] Whenever I write about significance in the following and if not indicated otherwise, he refers to statistical significance at conventional levels.

[77] The IV-Probit estimator obtains a p-value of 0.0448 for *Connection to Piped Water* respectively 0.0613 for *Connection to Improved Sewerage*.

[78] For the instrumented covariate *Connection to Piped Water* the Cragg-Donald Wald F-statistic is 36.78 compared to a Stock-Yogo critical reference value of 16.38, with the same conditions applying as above. For the covariate *Connection to Improved Sewerage* the F-statistic reads 29.90. The IV-Probit estimator obtains a p-value of 0.0121 for *Connection to Piped Water* respectively 0.0006 for *Connection to Improved Sewerage*.

[79] For the instrumented covariate *Connection to Piped Water* the Cragg-Donald Wald F-statistic is 10.47 compared to a Stock-Yogo critical reference value of 8.75, but only at a desired size of a Wald test of 0.20 (at a significance level of five per cent). For the covariate *Connection to Improved Sewerage* the F-statistic reads 8.08, compared to a reference value of 7.25 at a desired size of the Wald test of 0.25. 2SLS obtains a p-value of 0.0539 for *Connection to Piped Water,* and 0.0056 for *Connection to Improved Sewerage*, respectively regressed on two instruments.

[80] For the instrumented covariate *Connection to Piped Water* – and now three instead of two instruments – the Cragg-Donald Wald F-statistic is 91.19 compared to a Stock-Yogo critical reference value of 13.91 at a desired size of a Wald test of 0.05 (at a significance level of five per cent). For *Connection to Improved Sewerage,* the F-

Cleanser not much does change. Including only two of the three instruments leads to lower F-statistics, as before. With the full set of three instruments, though, validity is strongly supported.[81] The assumption of exogeneity still cannot be defended based on the Sargan-Hansen test, interestingly even less so when including the third instrument.[82]

When looking at the fifth and last outcome of interest – *Frequency of Latrine-Cleaning* – the pattern remains the same as with the two continuous expenditure variables before: validity can better be supported when including three instrumental variables, while the statistical evidence raising doubts about exogeneity is further diminished.[83]

The instruments chosen can be shown to be correlated with the decision to connect households to the piped water network respectively additionally to the sewerage network, while most likely not having an influence on probability of water treatment within the household or one of the other outcomes. I judge that none of the three instrumental variables can in any way be linked on grounds of rational argumentation with the outcome variables, as no transmission channels appear conceivable. Still, the difficulties to reject endogeneity of the instruments based on statistical tests create some concerns that the exclusion restriction might not be fulfilled. The results of the IV estimations therefore should be regarded with some caution.

While the instruments are employed in this analysis also with a linear-binary and a simultaneous maximum likelihood estimator (MLE) as well as with a two-stage Probit bootstrap (2SPrB) procedure,

---

statistic reads 83.71. 2SLS obtains a p-value of 0.0046 for *Connection to Piped Water* and 0.0036 for *Connection to Improved Sewerage,* respectively regressed on three instruments.

[81] For the instrumented covariate *Connection to Piped Water* the Cragg-Donald Wald F-statistic is 12.01 compared to a Stock-Yogo critical reference value of 11.59 at a desired size of a Wald test of 0.15 (at a significance level of five per cent), for two instruments. When including the third instrument the F-statistic rises to 84.27, which exceed the reference value of 13.91 for a Wald test size of 0.05. For *Connection to Improved Sewerage* and two instruments, the F-statistic reads 5.98, which is not even sufficient for a Wald test size of 0.25. When including the third instrument, the F-statistic rises to 77.59.

[82] 2SLS obtains a p-value of 0.0984 for *Connection to Piped Water* and 0.0237 for *Connection to Improved Sewerage,* respectively regressed on two instruments. When including the third instrument, the numbers change to 0.0012 and 0.0013, respectively.

[83] For the instrumented covariate *Connection to Piped Water* the Cragg-Donald Wald F-statistic is 9.98 compared to a Stock-Yogo critical reference value of 8.75 at a desired size of a Wald test of 0.20 (at a significance level of five per cent), for two instruments. When including the third instrument the F-statistic rises to 92.48, which exceed the reference value of 13.91 for a Wald test size of 0.05. For *Connection to Improved Sewerage* and two instruments, the F-statistic reads 6.76, which is not even sufficient for a Wald test size of 0.25. When including the third instrument, the F-statistic rises to 84.00. 2SLS obtains a p-value of 0.0995 for *Connection to Piped Water* and 0.07984for *Connection to Improved Sewerage,* respectively regressed on two instruments. When including the third instrument, the numbers change to 0.0021 and 0.0034, respectively.

the two-stage IV estimation shall serve as example to explain the mode of operation. For the first stage envision a linear probability model taking the form

$$\hat{C}_{ij} = \alpha_{i1} + \beta'_{i1}H'_j + \gamma'_{i1}E'_j + \delta'_{i1}Z'_j + \mu_{ij} \quad \text{(A.1)}$$

where $\hat{C}_{ij}$ indicates the predicted connection status *i* (piped water, sewerage) of each household *j* (C stands for Connection). The dependent variable is estimated based on a vector of characteristics of the head of the household (H stands for Household Head Characteristics), a vector of characteristics of the household itself as a whole (E stands for Environment) including location and ICT exposure measures, and a vector composed of the set of instrumental variables (Z stands for Instruments). The constant is indicated by α_1; β_1, $\gamma_1$and δ_1 stand for the respective coefficients of the vector variables of the first stage (subscript 1 stands for first stage). The error term μ_j varies non-systematically over the observation units so that I can write the conditional expectation function

$$E\left[\mu_j \mid H'_j, E'_j, Z'_j\right] = 0 \quad \text{(A.2)}$$

In the second stage, I plug the predicted connection status $\hat{C}_j$ estimated in equation (1) into the model estimating the dependent variable of interest. It takes the form

$$O_{dj} = \alpha_2 + \beta'_2 H'_j + \gamma'_2 E'_j + \delta_{i2}\,\hat{C}_{ij} + \varepsilon_j \quad \text{(A.3)}$$

with O_{dj} marking the reported behaviour of the household. O stands for Outcome; subscript *d* (which stands for Dependent Variable) indicates whether the outcome estimated is 1) probability of treatment of drinking water, 2) probability of using cleanser when washing hands, 3) logarithmised per capita spending on soap or 4) a combination of soap and laundry detergent, or 5) the frequency of latrine-cleaning. The other covariate vectors remain the same, the coefficients now marked with subscript 2 for the second stage. α_2 represents the constant, ε_j a non-systematic error term so that – equivalent to equation (2) – I can write the conditional expectation function as

$$E\left[\varepsilon_j \mid H'_j, E'_j, \hat{C}_{ij}\right] = 0 \quad \text{(A.4)}$$

The fact that both the dependent variable as well as the main explanatory variable are binary requires application of special estimators when trying to control for endogeneity. The choice of the appropriate estimator is not straightforward. In the literature one can find an array of methods applied. Each one comes with its assets and drawbacks, and no consensus on best practice has emerged yet. Holm and Arendt (2012) state that, in opposition to ”models with continuous endogenous repressors no simple consistent two-stage methods exist in the dummy endogenous model”. They point out that 2SLS and Probit models will yield diverging estimates if something else

than average effects are calculated, or in presence of outliers. Rees *et al.* (2001) use various methods as alternatives to a standard single-equation Probit estimator in order to allow for possible endogeneity issues. These include maximum likelihood bivariate Probit as well as 2SLS instrumental variable procedures. On the other hand, Angrist (2001) argues that nonlinear limited dependent variables (LDV, as for example a binary positive outcome variable) models might only *appear* to present particular difficulties regarding endogenous variables. He purports that causal inference in presence of LDV outcomes is similar to that of continuous outcomes. At least when there are no or "sparse and discrete" covariates, linear estimators are "no less appropriate for LDVs than for other kinds of dependent variables", he states.

In consideration of this lack of a widely accepted guideline I rely upon application of four estimation procedures to control for endogeneity. This allows me to confidently peg and delineate the conjectured range of coefficient size of the regressors. Besides a single-stage Probit estimator I deploy the following instrumental variable procedures in the analysis at hand: 1) a particular form of the bivariate Probit regression, 2) a two-stage Probit bootstrap regression, 3) a linear-binary regression (IV-Probit) combined with a range-check; and as an entirely linear procedure 4) two-stage least squares regression following a scheme as proposed by Wooldridge (2010). Below, these four procedures are considered one by one, with a focus on the less commonly used bivariate Probit estimator.

1) *Bivariate Probit:* Greene (2012) suggests that a model with an endogenous binary regressor can be treated analogously to a selection problem. The two-step procedure developed by Heckman (1978, see also Heckman, 1979, and Terza, 1986) is deemed inappropriate for application with a Probit model by Freedman and Sekhon (2010), as it is assessed to frequently compound bias. Also it is recognized to be inferior to maximum-likelihood procedures. They argue that under "ordinary circumstances, the two-step-correction should not be used in Probit response models" because it often increases bias. Carrasco (2001) chooses a bivariate Probit (otherwise also called Biprobit) model to avoid the "inappropriateness of the standard instrumental variable method for analysing the relationship between two endogenous discrete variables" in light of endogeneity. Chiburis, Das and Lokshin (2011) find that Biprobit outperforms two-stage instrumental estimators for sample sizes smaller than 5000, and when probability of treatment is close to the bounds of 0 and 1. They also demonstrate that the MLE procedure allows calculating Average Treatment Effects (ATE).

This motivates the application of a simultaneous estimation procedure, namely the bivariate Probit (Biprobit) estimator. The Biprobit procedure[84] subdivides into three options, which fulfil different functions. The regular bivariate Probit regression (BP) is according to Hardin (2012) suitable "for models where the two dependent variables depend on the same list of independent variables and are correlated". The Seemingly Unrelated Biprobit regression (SUR BP) is applied when the dependent variables of two equations "may not depend on the same list of independent variables, but are still correlated", while the nested Probit regression is used "where the outcome of one equation depends on the outcome of the other equation" (*ibid.*). The appropriate application in this study – in order to instrument a binary and potentially endogenous regressor in combination with a binary regressand – is a nested (or recursive, a label more commonly used in the literature) version of the SUR BP. I refer to it in the following as RSUR BP. Park (2010) shows how to correctly[85] formulate this RSUR BP model. Biprobit models assume that the error terms of the (simultaneously estimated) equations are independent and identically distributed (see Greene, 2012). Independence of the error terms and the instrumental variables is assumed.

The Biprobit estimators do not come without drawbacks, though. It is not appropriate when more than one regressor is endogenous, or when an endogenous regressor interacts with an exogenous one. The necessary assumptions regarding functional form and distributional considerations are strong compared to those of other estimators.

The application of the Biprobit estimator requires two assumptions to be fulfilled. First, the variance of the covariates must be sufficient (Dupas, 2011). The more restrictive assumption is the second: the errors are assumed to be "distributed bivariate standard normal" (Davis *et al.*, 2011); otherwise the identification by functional form is problematic. This assumption can be tested using the Murphy score test. In context of this study it turns out that the null hypotheses of bivariate standard normal distribution is rejected for all tested specifications except for two cases where the probability of drinking water treatment is regressed on connection to improved sewerage (test statistics are available on request). The null is rejected for all Biprobit specifications regressing the probability of cleanser use on any of the connection covariates; and for all those regressing the probability of treatment of drinking water on connection to piped water. I therefore refrain from considering the

[84] For details regarding the Biprobit estimator also refer to Maddala (1983) pp. 122-123 and Greene (2012), section 17.5, pp. 738 *et sqq*.

[85] The correct STATA coding scheme presents oneself as *biprobit (y = w x1-xn) (w = z x1-xn)*, with y denoting the binary outcome, x1-xn the exogenous regressors, w the potentially endogenous binary regressor, and z the instrumental variable or set of variables.

Biprobit to be a suitable estimator, but for purposes of documentation report the estimation results in Appendix C.6 (columns labelled SUR BP, in Table A.11, Table A.12, Table A.15, and Table A.16).

2) *Two-stage Probit bootstrap (2SPrB)*: an alternative estimation approach[86] employed is similar to that used by Noe and Rieckmann (2013). The difference is that I chose the Probit estimator for both stages in the study at hand; instead of employing Negative Binomial estimator for the first stage and a Probit estimator for the second. In the first stage the endogenous binary variable (henceforth abbreviated EBV in this study, here the connection dummy) is regressed on the exogenous covariates. In the second stage the outcome is regressed on the predicted value of the EBV and the exogenous covariates. I employ bootstrapping[87] to accurately calculate the standard errors, which otherwise would in tendency be understated.

3) *Linear-Binary (IV-Probit)*: I include the IV-Probit estimator for purpose of comparison, while verifying[88] upfront that the predicted value of the EBV regressed on instrumental variables and exogenous covariates is not far beyond the lower and upper bounds of 0 and 1. The maximum-likelihood option is preferred over the two-step option (*i.e.* the minimum chi-squared estimator introduced by Newey, 1987) as no convergence problems are encountered, and is superior in terms of efficiency. Note that some authors regard combining a linear model in the first stage with a nonlinear model in the second stage not to be appropriate when the endogenous explanatory variable is discrete (see, for example, Dong and Lewbel, 2012). Adkins (2009) point out (abbreviating IV-Probit to IVP) that the "bias of IVP is substantial when instruments are weak". Angrist and Pischke (2008) explain that the IV-Probit estimator can, but does not have to be, superior to a 2SLS approach.

4) *Wooldridge Procedure (W2SLS)*: the two-stage least squares instrumental variable procedure (see Turkington and Bowden, 1988, and Angrist and Krueger, 2001) may not be appropriate when the dependent variable is binary (see Newey, 1987). 2SLS allows to estimate "local average causal effects" (LATE, abbreviated alternatively ATT for Average Treatment effect on the Treated – see Angrist and Pischke, 2008), but may be biased and inferior to MLE when handling small samples. Wooldridge (2010)[89] nonetheless proposes a three-step IV procedure to estimate the Average

[86] This approach is based on an idea verbally suggested by Chris Müris.
[87] The number of iterations chosen is 10,000. This is way beyond the 500 iterations used, for instance, by Bernheim *et al.* (2001), and allows confident estimation of the standard errors.
[88] The two-stage IV-Probit procedure is here combined with an upstream range-check step using Ordinary Least Squares (OLS). This approach is based on a verbal suggestion given by Stephan Klasen.
[89] See procedure 21.1, pp. 939 *et seqq*.

Treatment Effect (ATE), combining a Probit with a 2SLS regression. He advocates first regressing the EBV on the set of exogenous covariates and the instrument, then in a second step predicting the value of the EBV. In a third step, he then regresses the outcome on the set of exogenous covariates and the EBV instrumented with its predicted value.

Appendix 5: Robustness checks in context of mapping hygiene habits

Spending on soap and laundry detergent

This section deals with the first two out of three[90] dependent variables that I deem to serve as alternative proxies of hygiene behaviour, thus supplementing *Probability of Cleanser Use*. These are 1) *Spending on Soap*, and 2) *Spending on Cleanser* (again referring to a combination of soap and detergent). Although there is no direct connection of expenses[91] for soap and laundry detergent with hygiene conditions within a household, the volume[92] of cleanser purchased allows deriving conclusions about which importance the household head attaches to matters of cleanliness of bodies and clothing. As the coefficients of the fitted models provide a very similar picture for both dependent variables, I discuss only the first one (*Spending on Soap*) here in detail. Regression tables for the second dependent variable *Spending on Cleanser* are displayed in the results Appendix 6 (see Table A.23 to Table A.24 and Table A.29). The role of a few explanatory variables slightly differs for the two dependent variables. These deviating cases I discuss below following the paragraph dealing with descriptive statistics.

In the mountainous region, mean spending on soap of households (see Table A.4 for descriptive statistics) is about six percentage points higher in the group connected to piped water than in the one additionally connected to improved sewerage, while both report between 13 (sewerage) to 19 (water) percentage points higher spending compared to the control group.

Table A.4: Mean spending (per capita, in YER) on soap by location and connection status

Connection	Mountain Project (YER)	N	SD	Mountain Control (YER)	N	SD	Coastal Project (YER)	N	SD	Coastal Control (YER)	N	SD
None	71.97	331	70.93	71.88	242	76.76	-	-	-	55.46	429	64.17
Piped Water	91.38	169	88.52	-	-	-	75.00	120	80.97	-	-	-
Sewerage	85.14	243	79.53	-	-	-	82.21	674	88.62	-	-	-

N over all connection groups and locations: 2208

Note: no households are connected to piped water or improved sewerage in the control towns. All households in the sample are connected in the coastal project town. Values capped at 500 YER per capita to exclude industrial use.

[90] The third alternative proxy is the dependent variable Frequency of Latrine Cleaning, addressed in the next section.

[91] Expenses may be driven by quantities purchased as well as by local prices. I assume that prices for soap as well as for detergent – each being products of quite uniform quality – do neither differ much within the project respectively control towns, nor between them. Therefore the level of expenses in a given household serves as indicator which quantity of the product was purchased, and which I assume to be a measure comparable over households from the whole sample.

[92] I expect households with more members to consume more cleanser than others. This is controlled for in several specifications.

In the coastal plain the positions of the two connected groups switch, with those connected to improved sewerage reporting to spend about seven percentage points more than the one connected to piped water only, and about 27 percentage points more than the control group. When combining spending on soap and laundry detergent, the group connected to improved sewerage consistently for both topographic regions reports higher spending than that connected to piped water, which in turn reports higher spending than the control group. There is one exception: in the mountains, the in-town control group reports lower spending than the households in the control town – see Table A.3 in Appendix 3.

To interpret the coefficients correctly, one must take into account that the dependent variable is logarithmised spending. This means the coefficients have to be exponentiated before they can be interpreted in terms of percentage changes.[93] For example the coefficients of the OLS estimator for the covariate connection to *Piped Water* ranges from 0.211 to 0.311, which corresponds to a percentage effect of between around 23.5 and 36.5 per cent for a one-unit change (which in this case is the step from 0, indicating no connection, and 1, indicating connection).

The overview table is structured as accustomed: I oppose the two covariates of main interest in specifications (1) to (6) in Table A.5 below. Table A.17 to Table A.20 in Appendix 6 display the estimated marginal effects for complementary specifications. As above I split the tables by main explanatory variables of interest, specifications vary media exposure. Results of alternative specifications including location, education and household composition variables, and the 2SLS estimator (Table A.19 and Table A.20) as well as results of first stages (see Table A.29) are presented in the same appendix. The coefficient of determination of the models ranges from seven to eleven per cent.

The amount of money spent by a household on soap is correlated with nine variables or groups of variables, for which I estimate coefficients statistically significant at conventional levels.

The most influential covariate with a positive sign of the coefficient is – following the interaction term of gender and education, which I discuss below – the *Region* a household is located in, with spending being roughly 40 per cent higher in the mountainous region compared to the coastal plain, keeping all other covariates stable. Estimated effect sizes I give as approximate percentages in

[93] A one-unit increase of the regressor must be interpreted differently for continuous and dummy regressors. For the former, the regressand in its non-logarithmised form changes by (coefficient size times 100) per cent. For the latter, the percental change of the regressand in its non-logarithmised form can be approximated by using the exponential function: [exp(coefficient size) minus one] times 100 per cent.

parentheses. In roughly descending order of estimated absolute coefficient sizes the other covariates with the same sign comprise: exposure to *ICT* (22.4 to 39.2%) with *Television* having the largest and *Web* access as the single exemption no effect at all, connection to *Piped Water* (23.5 to 36.5%), presence of a *Water Storage Tank* (19.9 to 28.5%), connection to *Improved Sewerage*[94] (16.2 to19.7%), *Assets* (8.6 to 16.1%), and *Symptom Occurrence* (10%).

Covariates with a negative effect are: the *Dependency Ratio* (the difference between the two extreme bounds – no dependent members in a household on the one hand, and the entire household being composed of them on the other – having an effect of between -44 and -56 per cent), location in the *Coastal Control* town (-32.2 to -47.8 per cent compared to the *Mountain Control* town), location in the *Coastal Project* town (-22.0 to -36.9%), and the *Age* of the household head (appearing to decrease expenditure by -0.4 to -0.6 per cent with each additional year of age).

No correlation at conventional levels of statistical significance I find for location in the *Mountainous Project* town compared to the *Mountainous Control* town, *Disease Awareness*, *Training* experience, or for characteristics of the household head as *Gender* and *Marital Status*. Interestingly, none of the alternative measures of education displays coefficients statistically significant at conventional levels – with one exception. The *Intermediate* (*i.e.*, between *Primary* and *Secondary*) *Level of Education* displays a negative coefficient (significant at the ten per cent level) ranging between -0.218 and -0.251 in those specifications which incorporate alternative household composition measures (*i.e.*, instead of the *Dependency Ratio* the *Number of Children*, *Elderly* or both). This corresponds to a change of -21.5 to -25.1 per cent compared to the reference category of no education. None of the other levels seems to have an influence.

When interacting[95] the *Gender* of the household head with the *Level of Education* the sign of intermediate education switches; and its sizable coefficient (1.013 to 1.036) becomes significant at the one per cent level. This corresponds to a percentage change of 175.4 to 181.8 per cent. The *Interaction Term* indicating *Gender* being male and education being *Intermediate* displays a negative coefficient, significant at the five per cent level, and ranging between -1.898 and -1.852. As *Gender* is an indicator variable taking the value 1 for male and 0 otherwise, the marginal effect of *Intermediate*

[94] Note that the explanatory variable connection to *Improved Sewerage* displays coefficients statistically significant at conventional levels in very few specifications.

[95] For reasons of space limitation, I do not display specifications including the interaction term. They are available upon request.

Table A.5: Spending on soap, connections compared, selected specifications

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	0.370***	0.357***	0.310***	0.311***	0.317***	0.320***
	(0.062)	(0.078)	(0.072)	(0.082)	(0.073)	(0.082)
Sewerage		0.019		-0.003		-0.006
		(0.083)		(0.086)		(0.086)
HH Member trained	0.024	0.022	0.019	0.019	0.029	0.029
	(0.140)	(0.141)	(0.140)	(0.141)	(0.140)	(0.141)
Web Access	-0.174	-0.175	-0.176	-0.176	-0.178	-0.178
	(0.327)	(0.327)	(0.326)	(0.326)	(0.323)	(0.323)
Primary Edu	0.112	0.112	0.109	0.109		
	(0.074)	(0.074)	(0.075)	(0.075)		
Intermediate Edu	-0.150	-0.151	-0.156	-0.156		
	(0.131)	(0.132)	(0.132)	(0.132)		
Secondary Edu	0.135	0.133	0.123	0.123		
	(0.095)	(0.095)	(0.095)	(0.095)		
Tertiary Edu	0.131	0.130	0.119	0.119		
	(0.108)	(0.108)	(0.108)	(0.108)		
Edu. Durat° HH Head					0.007	0.007
					(0.006)	(0.006)
Age HH Head	-0.004*	-0.004*	-0.005*	-0.005*	-0.004	-0.004
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Gender HH Head	-0.085	-0.084	-0.076	-0.076	-0.070	-0.070
	(0.183)	(0.184)	(0.183)	(0.184)	(0.182)	(0.182)
Marital Status HH Head	0.004	0.003	0.003	0.004	0.013	0.013
	(0.151)	(0.152)	(0.151)	(0.152)	(0.150)	(0.151)
Disease Awareness	0.068	0.068	0.065	0.065	0.062	0.062
	(0.056)	(0.056)	(0.057)	(0.057)	(0.057)	(0.057)
Asset Index	0.160***	0.159***	0.161***	0.161***	0.163***	0.163***
	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
Dependency Ratio	-0.551***	-0.549***	-0.534***	-0.535***	-0.535***	-0.536***
	(0.149)	(0.150)	(0.149)	(0.150)	(0.149)	(0.150)
Water Tank present	0.220**	0.220**	0.232***	0.232***	0.231***	0.231***
	(0.086)	(0.086)	(0.085)	(0.086)	(0.085)	(0.086)
Region	0.449***	0.452***				
	(0.067)	(0.069)				
Mountain Treatment			-0.095	-0.094	-0.096	-0.096
			(0.076)	(0.077)	(0.076)	(0.077)
Coastal Treatment			-0.446***	-0.445***	-0.449***	-0.447***
			(0.107)	(0.114)	(0.107)	(0.115)
Coastal Control			-0.601***	-0.601***	-0.603***	-0.602***
			(0.107)	(0.107)	(0.106)	(0.107)
Constant	3.228***	3.227***	3.755***	3.755***	3.729***	3.728***
	(0.200)	(0.201)	(0.212)	(0.212)	(0.211)	(0.211)
Sample Size	2262	2262	2262	2262	2262	2262
Adj. R^2	0.071	0.071	0.072	0.071	0.071	0.070

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

education in combination with the *Gender* being male has to be calculated[96] based on the sum of the education and interaction term coefficients (-0.816 to -0.885). This corresponds to a remarkable percentage change of -55.8 to -58.7 per cent. It cannot be entirely clarified why the household head being male and having completed an intermediate level of education appears more than halving the spending on soap compared with that of female household heads, ceteris paribus.

A similar picture I find for spending on cleanser when defined broader, as a combination of soap and detergent. An overview is given in Table A.21 below. Table A.21 to Table A.24 and Table A.29 in Appendix C.6 are complementary. A few changes arise: exposure to *ICT* gains in importance with estimated effects ranging between 25.7 per cent (*Landline* connection) to doubling (*Television Set*). *Training* comes in with an effect ranging between 42.2 and 54.0%. *Symptom Occurrence* is higher correlated (16 to 18%) and more often statistically significant at conventional levels. The coefficient of *Intermediate* education is now statistically significant at the ten per cent (in one case at the five per cent) level in more specifications, while the *Interaction Term* no longer is. Connection to *Piped Water* is significant only occasionally and at the one per cent level, while connection to *Improved Sewerage* appears to lose its importance altogether. Presence of a *Water Storage Tank* and possession of a *Radio* broadcasting receiver seem no longer to be influential.

As before, I test for robustness of the results by estimating different specifications. On top of that, I chose two dependent variables, defining it first narrow (soap only) and then broader (cleanser, *i.e.* soap and detergent). Within each of these two cases, the estimations are robust over all specifications for almost all of the explanatory variables regarding levels of statistical significance, direction of sign and relative coefficient size. In addition, over different specifications, the absolute coefficient sizes remain stable. Including alternative measures for education – *Duration* and *Literacy* instead of *Level* – and of household composition – *Number of Children*, *Elderly* or both – does not change the picture. Comparing the estimated coefficients of the explanatory variables for both dependent variables shows that they are of similar size and mostly display the same relative order. The main difference between the narrow and the broader defined models are the entry of *Training* experience and the exit of presence of a *Water Storage Tank* as supposable determinants.

[96] Note that for regressing the dependent variable Spending on Soap the OLS estimator is used. Thus the procedure introduced by Ai and Norton (2003); and used, for example, by Dreher *et al.* (2010) – analysing marginal effects of interaction terms in Probit regressions by means of graphs – does not apply here.

Table A.6: Spending on cleanser, connections compared, selected specifications

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	0.284***	0.233***	0.141*	0.125	0.149**	0.136
	(0.070)	(0.082)	(0.073)	(0.083)	(0.074)	(0.083)
Sewerage		0.073		0.030		0.025
		(0.084)		(0.086)		(0.086)
HH Member trained	0.440***	0.433***	0.430***	0.427***	0.438***	0.436***
	(0.104)	(0.105)	(0.104)	(0.105)	(0.104)	(0.105)
Web Access	-0.337	-0.336	-0.335	-0.335	-0.344	-0.344
	(0.363)	(0.364)	(0.362)	(0.362)	(0.359)	(0.359)
Primary Edu	0.048	0.047	0.044	0.044		
	(0.081)	(0.081)	(0.081)	(0.081)		
Intermediate Edu	-0.195	-0.201	-0.209	-0.211		
	(0.148)	(0.148)	(0.148)	(0.149)		
Secondary Edu	0.042	0.038	0.024	0.023		
	(0.102)	(0.102)	(0.103)	(0.103)		
Tertiary Edu	-0.008	-0.012	-0.027	-0.028		
	(0.119)	(0.119)	(0.120)	(0.120)		
Edu. Durat° HH Head					-0.000	-0.000
					(0.006)	(0.006)
Age HH Head	-0.009***	-0.009***	-0.009***	-0.009***	-0.009***	-0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Gender HH Head	-0.060	-0.055	-0.043	-0.041	-0.040	-0.039
	(0.190)	(0.191)	(0.190)	(0.191)	(0.190)	(0.190)
Marital Status HH Head	0.093	0.088	0.088	0.086	0.090	0.089
	(0.155)	(0.156)	(0.155)	(0.156)	(0.154)	(0.155)
Disease Awareness	0.261***	0.262***	0.258***	0.259***	0.258***	0.258***
	(0.061)	(0.061)	(0.062)	(0.061)	(0.062)	(0.061)
Asset Index	0.190***	0.189***	0.193***	0.192***	0.192***	0.192***
	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)
Dependency Ratio	-0.623***	-0.614***	-0.600***	-0.597***	-0.602***	-0.600***
	(0.165)	(0.166)	(0.165)	(0.166)	(0.166)	(0.166)
Water Tank present	0.071	0.069	0.088	0.086	0.085	0.084
	(0.092)	(0.092)	(0.091)	(0.092)	(0.091)	(0.092)
Region	0.506***	0.517***				
	(0.071)	(0.074)				
Mountain Treatment			-0.014	-0.017	-0.015	-0.017
			(0.073)	(0.074)	(0.074)	(0.075)
Coastal Treatment			-0.383***	-0.394***	-0.384***	-0.393***
			(0.108)	(0.115)	(0.109)	(0.115)
Coastal Control			-0.649***	-0.651***	-0.645***	-0.647***
			(0.115)	(0.116)	(0.115)	(0.115)
Constant	4.219***	4.215***	4.762***	4.766***	4.735***	4.739***
	(0.228)	(0.229)	(0.240)	(0.240)	(0.238)	(0.238)
Sample Size	1992	1992	1992	1992	1992	1992
Adj. R^2	0.077	0.077	0.078	0.078	0.078	0.078

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

I use 2SLS to control for endogeneity for the dependent variables referring to expenditures for soap and cleanser. Levels of statistical significance and coefficient signs are mostly consistent with all of the basic OLS models for most of the explanatory and control variables. The same issues apply as mentioned above and in Appendix 4 regarding the exogeneity assumption, so endogeneity – while I do not assume it to be likely – cannot be ruled out.

Frequency of latrine-cleaning

This section deals with the third alternative proxy of hygiene habits, employed to check for the robustness of the findings for the dependent variable *Probability of Cleanser Use*: the frequency, with which households use to clean their latrines. Descriptive statistics do not show striking differences between connection groups or topographic regions (see Table A.7).

Table A.7: Mean frequency of latrine-cleaning by location and connection status

Connection	Mountain Project	N	SD	Mountain Control	N	SD	Coastal Project	N	SD	Coastal Control	N	SD
None	7.86	355	3.28	7.77	291	2.64	-	-	-	7.19	401	3.30
Piped Water	8.12	197	3.57	-	-	-	7.65	124	2.92	-	-	-
Sewerage	7.80	268	3.39	-	-	-	8.60	713	3.28	-	-	-

N over all connection groups and locations: 2349

Note: no households are connected to piped water or improved sewerage in the control towns. All households in the sample are connected in the coastal project town.

Households tend to report the frequency of cleaning to be daily. Frequency is reported in times per week, and the modus is seven. This may either reflect a widespread habit to actually conduct cleaning daily, or the Hawthorne effect mentioned above in the data section.

An overview is given in Table A.8 below. Table A.25 to Table A.28 and Table A.29 in the results appendix are complementary. The coefficient of determination of the models ranges from three to five per cent.

The frequency of latrine cleaning is positively correlated with nine explanatory variables or groups of variables, which are endowed with coefficients statistically significant at conventional levels. Coefficient size ranges I show in parentheses. Due to the employment of a linear estimator, I can interpret them straightforward as percentage effects. For instance, a coefficient sized 0.834 corresponds to an increase of frequency of latrine cleaning of about 83 per cent, ceteris paribus.

Table A.8: Latrine-cleaning frequency, connections compared, selected specifications

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	0.757***	0.606***	0.478**	0.420*	0.476**	0.419*
	(0.141)	(0.208)	(0.219)	(0.254)	(0.217)	(0.252)
Sewerage		0.232		0.125		0.123
		(0.212)		(0.212)		(0.212)
HH Member trained	0.987**	0.962**	0.967**	0.956**	0.959**	0.947**
	(0.468)	(0.466)	(0.469)	(0.468)	(0.469)	(0.467)
Web Access	-0.052	-0.061	-0.055	-0.060	-0.006	-0.011
	(0.839)	(0.840)	(0.843)	(0.843)	(0.849)	(0.849)
Primary Edu	-0.049	-0.051	-0.060	-0.060		
	(0.179)	(0.179)	(0.179)	(0.179)		
Intermediate Edu	-0.081	-0.094	-0.115	-0.119		
	(0.267)	(0.268)	(0.270)	(0.270)		
Secondary Edu	-0.142	-0.156	-0.189	-0.194		
	(0.209)	(0.210)	(0.210)	(0.211)		
Tertiary Edu	0.330	0.317	0.278	0.275		
	(0.259)	(0.258)	(0.260)	(0.260)		
Edu. Durat° HH Head					-0.001	-0.002
					(0.014)	(0.014)
Age HH Head	0.004	0.004	0.003	0.003	0.004	0.004
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Gender HH Head	0.082	0.093	0.125	0.129	0.117	0.121
	(0.303)	(0.303)	(0.302)	(0.302)	(0.304)	(0.304)
Marital Status HH Head	0.144	0.134	0.143	0.137	0.148	0.143
	(0.250)	(0.249)	(0.250)	(0.250)	(0.250)	(0.250)
Disease Awareness	0.003	0.005	0.004	0.005	0.003	0.004
	(0.135)	(0.135)	(0.134)	(0.134)	(0.134)	(0.134)
Asset Index	0.604***	0.600***	0.614***	0.611***	0.622***	0.619***
	(0.106)	(0.106)	(0.106)	(0.106)	(0.106)	(0.106)
Dependency Ratio	0.253	0.282	0.321	0.335	0.343	0.356
	(0.303)	(0.304)	(0.303)	(0.304)	(0.305)	(0.306)
Water Tank present	0.499***	0.492***	0.553***	0.547***	0.569***	0.562***
	(0.173)	(0.173)	(0.174)	(0.174)	(0.173)	(0.173)
Region	-0.237	-0.197				
	(0.161)	(0.161)				
Mountain Treatment			-0.292	-0.309	-0.293	-0.310
			(0.243)	(0.241)	(0.242)	(0.240)
Coastal Treatment			0.360	0.305	0.375	0.320
			(0.316)	(0.315)	(0.318)	(0.317)
Coastal Control			-0.380	-0.390	-0.370	-0.381
			(0.240)	(0.239)	(0.239)	(0.239)
Constant	5.543***	5.518***	5.529***	5.545***	5.456***	5.471***
	(0.443)	(0.442)	(0.479)	(0.479)	(0.475)	(0.474)
Sample Size	2407	2407	2407	2407	2407	2407
Adj. R^2	0.041	0.041	0.044	0.044	0.044	0.044

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

In roughly descending order of estimated coefficient size the covariates positively correlated are *Mean Education Duration* of adults in a household (strongly correlated with a coefficient of 1.071 to 1.115, but only when the covariate *Dependency Ratio* is included in the specification), *Training* experience (0.834 to 0.966), and the *Dependency Ratio* (0.663 to 0.667, only when *Mean Education Duration* of adults is included).

The next in line are connection to *Piped Water* (0.380 to 0.721), wealth reflected by the *Asset Index* (0.544 to 0.611, significant at the one per cent level), and location in the *Coastal Project* town (0.506, compared to reference category *Mountain Control* town, significant only in combination with connection to *Piped Water*, but not to *Improved Sewerage*).

Then follow connection to *Improved Sewerage* (0.581, only in combination with *Region*), presence of a *Water Storage Tank* (0.464 to 0.559, significant at the one per cent level), *Tertiary Level of Education* (0.453, compared to reference category no education, only when the covariate *Region* is included), and the three alternative measures of household composition (*Number of Household Members* with 0.046 to 0.047, *Number of Children* with 0.055, *Number of Children and Elderly* with 0.053 to 0.054).

On the other hand, *Frequency of Latrine-Cleaning* is negatively correlated with *Primary* (-0.857 to -0.861) and *Secondary Education* when the interaction term (*Level of Education* interacted with *Gender* of the household head, itself not significant) is included in the specification. Location in the *Coastal Control* town is negatively correlated, likewise (-0.425 to -0.449), though not significant at conventional levels in all but two specifications. Location in the *Mountainous Region* is statistically significant at the ten per cent level in merely one specification, with a coefficient size of -0.299; and on top is unstable regarding sign.

Obviously beside the point are some of the other covariates. These include *Age*, *Gender* and *Marital Status* of the household head, *Disease Awareness*, and location in the *Mountain Project* town, an indicator whether any *Soap* was bought[97] within a recall period of 30 days, exposure to *ICT* and – as briefly mentioned above – the *Interaction Term*.

The results I test for robustness following the same scheme as before, *i.e.* estimating different specifications. Coefficient signs, sizes and levels of statistical significance are largely consistent. Because biased (over-)reporting due to a Hawthorne cannot be dismissed, the findings of this

[97] This might give a clue about the importance attached by household members to cleanliness.

section should not be overrated, though. Nevertheless, they can be put into perspective with the findings for the antecedent hygiene habit regressands, and contribute to a more complete picture. 2SLS specifications reveal no critical indication of endogeneity issues, but are of limited informative value due to the concerns regarding exogeneity of the instruments.

In the following, I discuss the patterns of likely determinants of hygiene habits found for the four dependent variables (probability of *Use of Cleanser* for hand washing, *Spending on Soap* and on *Cleanser*, *Frequency of Latrine-Cleaning*) used to proxy those hygiene habits. The advance is analogical to the structure used above in the discussion of the findings for the dependent variable *Probability of Drinking Water Treatment*. First, I recall each of the theories laid down in the introduction. Then I assess the role of the respective covariate or group of covariates as a – potentially joint – determinant of the group of four dependent variables, with attention to sign and rank. Covariates whose correlation coefficients are never statistically significant at least at the ten per cent level I do not regard to be influential determinants.

Hypothesis 1b (connecting a household to piped water increases its likelihood of displaying hygiene habits) is consistently supported by the findings for each of the dependent variables. Connection to *Piped Water* ranks among the covariates with the most sizable influence, usually between third and fifth position. *Hypothesis 2b* (additional connection of a household to *Improved Sewerage* decreases its likelihood of displaying hygiene habits, but to a lesser extent than connection to piped water) can be supported as well. The correlation is positive for three out of four dependent variables, solely with *Spending on Cleanser* apparently not being determined by connection to *Improved Sewerage*. The covariate ranks lower than connection to *Piped Water* for all outcomes. *Hypothesis 3b* (those households in which at least one member participated in any type of hygiene *Training* in the past have a higher likelihood of displaying hygiene habits) is distinctly reinforced, with coefficients ranking between first and fifth position. Interestingly, *Training* experience appears to be the most important covariate for the outcome *Spending on Cleanser* including laundry detergent, while it seems to be irrelevant for *Spending on Soap*. *Hypothesis 4b* (those households with exposure to information and communication technology have a higher likelihood of displaying hygiene habits) is consistently supported, with the exception of the outcome *Frequency of Latrine-Cleaning*. For the latter, the coefficients never reach the ten per cent level of statistical significance. Please note that there is, in contrast to expectation, no evidence that access to the *World Wide Web* is a driver. Instead, those *ICT* measures that indicate possession of objects (*Television Set*, *Personal Computer*, *Satellite Dish*, *Radio* broadcasting receiver and *Mobile Phone*) make the impression to be. This gives room for doubt if the correlation does indeed map the influence of increased availability of

information, or rather depicts the wealth of the household. On the other hand, the correlation of connection to a phone *Landline* with the outcome variables softens this concern slightly, as I believe such a connection to be far less costly than purchase of electronic goods. Another source of reassurance is the fact that the *Asset Index* is used to control for wealth.

Hypothesis 5b (the more educated the head of the household is, the higher is the likelihood of the household members displaying hygiene habits) can neither be supported nor invalidated. Empirical evidence is mixed. It indicates a positive relationship of *Tertiary* (for the *Use of Cleanser* also of *Secondary*) *Education* on *Use of Cleanser* for hand washing on the one hand, and on *Frequency of Latrine-Cleaning* on the other. However, regarding the expenditure outcomes, no effect is revealed by the analysis, except for *Intermediate* education – the coefficients of which are negatively correlated with the outcomes.

The evidence regarding the control variables covering household composition is mixed. While the probability of *Use of Cleanser* and the *Spending on Soap* seem to be negatively and – with the coefficient in most specifications being the second largest – strongly influenced, the contrary appears to be the case for *Spending on Cleanser* and *Latrine-Cleaning Frequency*. Female headship does not seem to play an influential role. The age of the household head is also not fostering hygiene habits. For two of the outcomes, namely *Use of Cleanser* for hand washing and *Frequency of Latrine-Cleaning*, the coefficients never reach the ten per cent level of statistical significance. For both expenditure outcomes, a negative correlation – statistically significant at conventional levels up to the one per cent level – is estimated for most specifications. This is in line with my expectation of younger household heads being more prone to pursue hygiene habits, and contradicts the previous literature. Considering that income tends to increase with *Age*, the relationship between young age and spending might indeed be understated here, if income is not fully controlled for by the asset index. This latter control variable is consistently positively correlated with all dependent variables proxying hygiene habits. The coefficient usually ranks between sixth and ninth position in terms of relative size. At least one household member having suffered from water-related diseases in the recent past (*i.e.*, *Illness History*) is positively correlated with the likelihood of displaying hygiene habits. Likewise, the covariate *Symptom Occurrence* is consistently positively correlated with the outcome variables.

Appendix 6: Estimation results

Table A.9: Probability of drinking water treatment, piped water, ICT covariates

Specification	(1)	(2)	(3)	(4)	(5)
Estimator	Probit	Probit	Probit	Probit	Probit
Piped Water	0.055***	0.053***	0.054***	0.053***	0.052***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Age HH Head	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Gender HH Head	-0.045*	-0.044*	-0.044*	-0.050**	-0.044*
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Marital Status HH Head	0.013	0.012	0.013	0.011	0.012
	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)
Disease Awareness	0.007	0.006	0.007	0.005	0.005
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
HH Member trained	0.104***	0.104***	0.102***	0.102***	0.106***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Asset Index	0.040***	0.038***	0.038***	0.034***	0.036***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)
Dependency Ratio	0.000	0.000	0.005	0.005	0.006
	(0.028)	(0.028)	(0.027)	(0.028)	(0.028)
Region	0.153***	0.153***	0.146***	0.149***	0.152***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Water Tank present	-0.021	-0.022	-0.022	-0.021	-0.021
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
TV present	0.009				
	(0.019)				
Sat. Dish present		0.019			
		(0.013)			
Radio present			0.025**		
			(0.011)		
GSM present				0.032**	
				(0.014)	
Landline Connection					0.028**
					(0.012)
Primary Edu	0.045***	0.044***	0.044***	0.045***	0.044***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Intermediate Edu	0.011	0.009	0.009	0.012	0.010
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Secondary Edu	0.045**	0.043**	0.043**	0.044**	0.042**
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Tertiary Edu	0.046**	0.044**	0.042**	0.044**	0.043**
	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)
Sample Size	2459	2459	2459	2459	2459
Pseudo R^2	0.173	0.174	0.176	0.176	0.176

Average marginal effects displayed. Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.10: Probability of drinking water treatment, sewerage, ICT covariates

Specification	(1)	(2)	(3)	(4)	(5)
Estimator	Probit	Probit	Probit	Probit	Probit
Piped Water	0.033**	0.032**	0.032**	0.032**	0.031**
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Sewerage	0.037**	0.036**	0.038**	0.037**	0.037**
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Age HH Head	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Gender HH Head	-0.043*	-0.042*	-0.042*	-0.047*	-0.042*
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Marital Status HH Head	0.011	0.010	0.011	0.009	0.010
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Disease Awareness	0.007	0.007	0.007	0.005	0.006
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
HH Member trained	0.099***	0.098***	0.096***	0.097***	0.100***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Asset Index	0.039***	0.038***	0.037***	0.034***	0.036***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)
Dependency Ratio	0.004	0.004	0.009	0.009	0.010
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Region	0.160***	0.160***	0.153***	0.156***	0.159***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Water Tank present	-0.021	-0.021	-0.022	-0.020	-0.020
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
TV present	0.008				
	(0.019)				
Sat. Dish present		0.017			
		(0.013)			
Radio present			0.025**		
			(0.011)		
GSM present				0.031**	
				(0.014)	
Landline Connection					0.027**
					(0.012)
Primary Edu	0.044***	0.044***	0.044***	0.044***	0.044***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Intermediate Edu	0.007	0.007	0.007	0.007	0.007
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Secondary Edu	0.040**	0.040**	0.040**	0.040**	0.040**
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Tertiary Edu	0.042**	0.042**	0.042**	0.042**	0.042**
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Sample Size	2459	2459	2459	2459	2459
Pseudo R^2	0.177	0.178	0.180	0.180	0.180

Average marginal effects displayed. Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.11: Probability of drinking water treatment, piped water, other specifications & estimators

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	Probit	Probit	IV-Probit	SUR BP	2SLS	2SPrB
Piped Water	0.061***	0.060***	0.966***	0.054***	0.525***	0.092
	(0.013)	(0.014)	(0.286)	(0.017)	(0.172)	(0.694)
Age HH Head	-0.000	-0.001*	-0.002	-0.000	-0.002*	0.000
	(0.000)	(0.000)	(0.003)	(0.000)	(0.001)	(0.005)
Gender HH Head	-0.034	-0.022	-0.325*	-0.020*	-0.021	-0.054*
	(0.023)	(0.023)	(0.175)	(0.010)	(0.031)	(0.207)
Marital Status HH Head	0.007	0.013	0.138	0.005	0.022	0.016
	(0.020)	(0.020)	(0.151)	(0.008)	(0.027)	(0.174)
Disease Awareness	0.005	0.003	0.082	0.002	0.025	0.009
	(0.011)	(0.011)	(0.080)	(0.004)	(0.016)	(0.092)
Symptom Occurrence			0.064	0.004	-0.006	0.008
			(0.084)	(0.005)	(0.017)	(0.014)
HH Member trained	0.107***	0.103***	0.694***	0.042***	0.118***	0.116***
	(0.021)	(0.021)	(0.164)	(0.012)	(0.040)	(0.188)
Asset Index	0.039***	0.037***	0.281***	0.016***	0.035***	0.038***
	(0.007)	(0.007)	(0.053)	(0.004)	(0.011)	(0.057)
Dependency Ratio	-0.003	0.030	-0.007	-0.004	0.045	-0.004
	(0.027)	(0.029)	(0.205)	(0.011)	(0.036)	(0.242)
Mountain Treatment	0.035**	0.035**	0.005	0.014**	-0.115	0.031
	(0.016)	(0.016)	(0.178)	(0.007)	(0.073)	(0.305)
Coastal Treatment	-0.146***	-0.153***	-1.520***	-0.062***	-0.508***	-0.192**
	(0.024)	(0.024)	(0.273)	(0.014)	(0.142)	(0.571)
Coastal Control	-0.072***	-0.073***	-0.492***	-0.030***		
	(0.024)	(0.024)	(0.175)	(0.011)		
Water Tank present	-0.025	-0.028	-0.221	-0.010	-0.058***	-0.036
	(0.018)	(0.018)	(0.135)	(0.008)	(0.018)	(0.172)
TV and/or PC present						0.009
						(0.183)
Web Access	0.072**	0.077**	0.470*	0.030**	0.107*	
	(0.032)	(0.033)	(0.243)	(0.015)	(0.063)	
Edu. Durat° HH Head					-0.001	
					(0.002)	
Literacy HH Head	0.034**					
	(0.014)					
Mean Edu. Durat° Adults		0.095***				
		(0.030)				
Primary Edu			0.283***	0.018***		0.046**
			(0.110)	(0.006)		(0.129)
Intermediate Edu			0.062	0.004		0.009
			(0.193)	(0.009)		(0.224)
Secondary Edu			0.276**	0.020**		0.057**
			(0.132)	(0.008)		(0.180)
Tertiary Edu			0.242	0.020**		0.058**
			(0.150)	(0.009)		(0.192)
Sample Size	2459	2464	2393	2459	2029	2029
Pseudo R^2	0.180	0.185				0.157

Average marginal effects displayed for non-linear models. Robust standard errors in parentheses; bootstrapped for 2SPrB. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.12: Probability of drinking water treatment, sewerage, other specifications & estimators

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	Probit	Probit	IV-Probit	SUR BP	2SLS	2SPrB
Piped Water	0.043***	0.042***	0.776*	0.044**	0.251***	-0.003
	(0.015)	(0.015)	(0.464)	(0.019)	(0.077)	(0.755)
Sewerage	0.043***	0.041***	0.096	0.017**	-0.019	0.072
	(0.015)	(0.016)	(0.240)	(0.007)	(0.028)	(0.420)
Age HH Head	-0.000	-0.001*	-0.002	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.004)	(0.000)	(0.001)	(0.005)
Gender HH Head	-0.033	-0.021	-0.336*	-0.019*	-0.032	-0.053*
	(0.023)	(0.023)	(0.177)	(0.011)	(0.025)	(0.210)
Marital Status HH Head	0.007	0.013	0.137	0.004	0.018	0.010
	(0.020)	(0.020)	(0.153)	(0.009)	(0.022)	(0.183)
Disease Awareness	0.006	0.005	0.076	0.003	0.012	0.006
	(0.011)	(0.011)	(0.080)	(0.004)	(0.013)	(0.089)
Symptom Occurrence			0.063	0.004	0.003	0.011
			(0.086)	(0.005)	(0.015)	(0.093)
HH Member trained	0.104***	0.099***	0.704***	0.041***	0.149***	0.115***
	(0.021)	(0.021)	(0.165)	(0.012)	(0.039)	(0.169)
Asset Index	0.040***	0.038***	0.283***	0.015***	0.039***	0.037***
	(0.007)	(0.007)	(0.053)	(0.004)	(0.010)	(0.056)
Dependency Ratio	0.001	0.033	-0.018	-0.002	0.008	-0.009
	(0.027)	(0.029)	(0.207)	(0.011)	(0.030)	(0.230)
Mountain Treatment	0.026	0.026	0.044	0.011	0.002	0.041
	(0.017)	(0.017)	(0.183)	(0.007)	(0.030)	(0.214)
Coastal Treatment	-0.166***	-0.170***	-1.464***	-0.071***	-0.273***	-0.175***
	(0.025)	(0.025)	(0.260)	(0.017)	(0.050)	(0.302)
Coastal Control	-0.075***	-0.076***	-0.514***	-0.032***		
	(0.024)	(0.024)	(0.184)	(0.012)		
Water Tank present	-0.026	-0.028	-0.210	-0.010	-0.039***	-0.030
	(0.018)	(0.018)	(0.135)	(0.008)	(0.015)	(0.159)
TV and/or PC present						0.018
						(0.176)
Web Access			0.470*	0.028*	0.140**	
			(0.251)	(0.015)	(0.064)	
Edu. Durat° HH Head					0.002	
					(0.001)	
Literacy HH Head	0.034**					
	(0.014)					
Mean Edu. Durat° Adults		0.090***				
		(0.030)				
Primary Edu			0.296***	0.018***		0.050***
			(0.112)	(0.007)		(0.133)
Intermediate Edu			0.063	0.004		0.009
			(0.195)	(0.009)		(0.227)
Secondary Edu			0.304**	0.020**		0.067***
			(0.135)	(0.008)		(0.169)
Tertiary Edu			0.272*	0.020**		0.068**
			(0.154)	(0.009)		(0.190)
Sample Size	2459	2463	2393	2459	2029	2029
Pseudo R^2	0.183	0.185				0.172

Average marginal effects displayed for non-linear models. Robust standard errors in parentheses; bootstrapped for 2SPrB . Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.13: Probability of cleanser use, piped water, ICT covariates

Specification	(1)	(2)	(3)	(4)	(5)
Estimator	Probit	Probit	Probit	Probit	Probit
Piped Water	0.121***	0.118***	0.122***	0.120***	0.116***
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Age HH Head	0.000	0.000	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gender HH Head	-0.001	-0.001	0.001	-0.006	-0.000
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
Marital Status HH Head	-0.001	-0.000	0.000	-0.003	-0.002
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Disease Awareness	0.040**	0.038**	0.041**	0.038**	0.040**
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
HH Member trained	0.091*	0.092*	0.089*	0.090*	0.098**
	(0.050)	(0.050)	(0.050)	(0.050)	(0.049)
Asset Index	0.040***	0.038***	0.039***	0.031**	0.032**
	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)
Dependency Ratio	-0.195***	-0.197***	-0.190***	-0.189***	-0.183***
	(0.041)	(0.041)	(0.041)	(0.040)	(0.040)
Mountain Treatment	0.003	0.002	-0.001	-0.001	0.001
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Coastal Treatment	0.157***	0.153***	0.163***	0.161***	0.156***
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
Coastal Control	0.224***	0.219***	0.223***	0.220***	0.219***
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
Water Tank present	0.008	0.009	0.010	0.013	0.009
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
TV present	0.043*				
	(0.025)				
Sat. Dish present		0.048**			
		(0.019)			
Radio present			0.044**		
			(0.019)		
GSM present				0.048**	
				(0.020)	
Landline Connection					0.087***
					(0.022)
Primary Edu	-0.006	-0.008	-0.006	-0.005	-0.007
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Intermediate Edu	0.052	0.051	0.053	0.054	0.050
	(0.039)	(0.039)	(0.039)	(0.038)	(0.038)
Secondary Edu	0.082***	0.079***	0.082***	0.081***	0.077***
	(0.028)	(0.029)	(0.028)	(0.028)	(0.028)
Tertiary Edu	0.098***	0.094***	0.096***	0.096***	0.091***
	(0.031)	(0.032)	(0.032)	(0.032)	(0.032)
Sample Size	2466	2466	2466	2466	2466
Pseudo R^2	0.090	0.091	0.090	0.090	0.094

Average marginal effects displayed. Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.14: Probability of cleanser use, sewerage, ICT covariates

Specification	(1)	(2)	(3)	(4)	(5)
Estimator	Probit	Probit	Probit	Probit	Probit
Piped Water	0.099***	0.096***	0.100***	0.098***	0.095***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Sewerage	0.048*	0.047*	0.049*	0.049*	0.047*
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Age HH Head	0.000	0.000	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gender HH Head	0.001	0.001	0.002	-0.005	0.001
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
Marital Status HH Head	-0.004	-0.002	-0.002	-0.005	-0.005
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Disease Awareness	0.041**	0.038**	0.042**	0.038**	0.040**
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
HH Member trained	0.084*	0.086*	0.083*	0.084*	0.091*
	(0.050)	(0.050)	(0.050)	(0.050)	(0.049)
Asset Index	0.039***	0.037***	0.038***	0.030**	0.031**
	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)
Dependency Ratio	-0.189***	-0.191***	-0.184***	-0.183***	-0.178***
	(0.041)	(0.040)	(0.041)	(0.040)	(0.040)
Mountain Treatment	-0.004	-0.004	-0.007	-0.007	-0.005
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Coastal Treatment	0.137***	0.134***	0.142***	0.140***	0.137***
	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
Coastal Control	0.219***	0.215***	0.219***	0.216***	0.215***
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
Water Tank present	0.006	0.007	0.007	0.010	0.007
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
TV present	0.042*				
	(0.025)				
Sat. Dish present		0.047**			
		(0.019)			
Radio present			0.043**		
			(0.019)		
GSM present				0.048**	
				(0.020)	
Landline Connection					0.086***
					(0.022)
Primary Edu	-0.007	-0.008	-0.006	-0.006	-0.008
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Intermediate Edu	0.050	0.049	0.051	0.052	0.048
	(0.039)	(0.039)	(0.039)	(0.038)	(0.038)
Secondary Edu	0.080***	0.077***	0.080***	0.079***	0.075***
	(0.029)	(0.029)	(0.028)	(0.028)	(0.028)
Tertiary Edu	0.097***	0.093***	0.094***	0.094***	0.089***
	(0.031)	(0.032)	(0.032)	(0.032)	(0.032)
Sample Size	2466	2466	2466	2466	2466
Pseudo R^2	0.091	0.092	0.091	0.092	0.095

Average marginal effects displayed. Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.15: Probability of cleanser use, piped water, other specifications & estimators

Specification Estimator	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) Probit	(6) IV-Probit	(7) SUR BP	(8) 2SLS	(9) 2SPrB
Piped Water	0.116***	0.124***	0.123***	0.130***	0.125***	0.877***	0.218***	0.292	-0.082
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.203)	(0.045)	(0.193)	(0.557)
Age HH Head	0.001	-0.000	-0.001	0.001	0.000	0.000	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.000)	(0.001)	(0.003)
Gender HH Head	-0.015	0.019	0.033	0.028	0.016	0.012	-0.002	0.018	0.018
	(0.046)	(0.046)	(0.045)	(0.045)	(0.046)	(0.148)	(0.022)	(0.046)	(0.174)
Marital Status HH Head	-0.003	-0.019	-0.013	-0.021	-0.003	0.011	0.002	-0.048	-0.063
	(0.040)	(0.041)	(0.040)	(0.039)	(0.040)	(0.130)	(0.019)	(0.039)	(0.153)
Disease Awareness	0.037**	0.041**	0.040**	0.040**	0.040**	0.129**	0.021**	0.067***	0.050**
	(0.017)	(0.018)	(0.017)	(0.017)	(0.017)	(0.057)	(0.009)	(0.021)	(0.070)
Symptom Occurrence						0.052	0.013	0.034	0.045**
						(0.062)	(0.010)	(0.022)	(0.072)
HH Member trained	0.087*	0.093*	0.091*	0.100**	0.098**	0.249	0.043*	0.043	0.115**
	(0.049)	(0.050)	(0.050)	(0.050)	(0.050)	(0.166)	(0.026)	(0.038)	(0.180)
Asset Index	0.037***	0.044***	0.043***	0.048***	0.047***	0.130***	0.023***	0.039***	0.045***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.042)	(0.007)	(0.013)	(0.048)
Dependency Ratio	-0.202***	-0.179***	-0.151***			-0.596***	-0.097***	-0.205***	-0.248***
	(0.041)	(0.040)	(0.042)			(0.135)	(0.027)	(0.050)	(0.170)
# of HH Members				-0.007***					
				(0.002)					
# of Children					-0.019***				
					(0.004)				
Mountain Treatment	0.004	0.003	0.003	-0.002	-0.001	-0.176	0.001	-0.057	0.078
	(0.027)	(0.028)	(0.027)	(0.027)	(0.027)	(0.117)	(0.013)	(0.084)	(0.239)
Coastal Treatment	0.155***	0.175***	0.170***	0.161***	0.155***	0.098	0.073***	0.023	0.329**
	(0.035)	(0.035)	(0.035)	(0.036)	(0.036)	(0.198)	(0.021)	(0.157)	(0.472)
Coastal Control	0.224***	0.233***	0.231***	0.225***	0.222***	0.742***	0.099***		
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.103)	(0.026)		
Water Tank present	0.006	0.015	0.014	0.017	0.014	0.022	0.008	0.002	0.039
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.078)	(0.012)	(0.027)	(0.106)
TV present	0.041*	0.044*	0.040	0.050**	0.048*				
	(0.025)	(0.025)	(0.025)	(0.025)	(0.024)				
TV and/or PC present									0.043
									(0.113)
Web Access	0.032	0.041	0.044	0.030	0.032	0.045	0.016	-0.008	
	(0.082)	(0.082)	(0.082)	(0.082)	(0.083)	(0.268)	(0.041)	(0.071)	
Edu. Durat° HH Head	0.009***							0.008***	
	(0.002)							(0.003)	
Literacy HH Head		0.045**							
		(0.021)							
Mean Edu. Durat° Adults			0.091*						
			(0.051)						
Primary Edu				-0.008	-0.009	-0.023	-0.003		0.033
				(0.024)	(0.024)	(0.074)	(0.011)		(0.092)
Intermediate Edu				0.041	0.043	0.168	0.026		0.071
				(0.039)	(0.039)	(0.127)	(0.020)		(0.145)
Secondary Edu				0.065**	0.070**	0.224**	0.040**		0.134***
				(0.029)	(0.028)	(0.098)	(0.016)		(0.134)
Tertiary Edu				0.079**	0.084***	0.274**	0.050***		0.137***
				(0.032)	(0.031)	(0.111)	(0.018)		(0.146)
Sample Size	2466	2466	2471	2466	2466	2400	2466	2037	2037
Adj. R^2								0.105	
Pseudo R^2	0.093	0.086	0.085	0.085	0.089				0.097

Average marginal effects displayed for non-linear models. Robust standard errors in parentheses; bootstrapped for 2SPrB. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.16: Probability of cleanser use, sewerage, other specifications & estimators

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimator	Probit	Probit	Probit	Probit	Probit	IV-Probit	SUR BP	2SLS	2SPrB
Piped Water	0.094***	0.100***	0.100***	0.105***	0.103***	0.911***	0.203***	0.227**	0.047
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.329)	(0.050)	(0.111)	(0.338)
Sewerage	0.047*	0.053*	0.051*	0.054**	0.048*	-0.055	0.024*	0.002	0.070
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.163)	(0.014)	(0.050)	(0.162)
Age HH Head	0.001	-0.000	-0.001	0.001	0.000	0.000	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.000)	(0.001)	(0.003)
Gender HH Head	-0.014	0.020	0.033	0.029	0.017	0.009	-0.001	0.016	0.026
	(0.046)	(0.046)	(0.045)	(0.045)	(0.045)	(0.147)	(0.023)	(0.045)	(0.167)
Marital Status HH Head	-0.005	-0.020	-0.015	-0.022	-0.005	0.013	0.001	-0.050	-0.061
	(0.040)	(0.041)	(0.040)	(0.039)	(0.040)	(0.131)	(0.020)	(0.039)	(0.147)
Disease Awareness	0.038**	0.042**	0.041**	0.040**	0.040**	0.129**	0.021**	0.064***	0.058***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.057)	(0.009)	(0.020)	(0.064)
Symptom Occurrence						0.050	0.013	0.037*	0.041*
						(0.063)	(0.010)	(0.021)	(0.069)
HH Member trained	0.081	0.086*	0.082	0.092*	0.091*	0.247	0.040	0.050	0.082
	(0.049)	(0.050)	(0.051)	(0.050)	(0.050)	(0.165)	(0.026)	(0.035)	(0.167)
Asset Index	0.036***	0.043***	0.042***	0.047***	0.046***	0.131***	0.022***	0.040***	0.042***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.042)	(0.007)	(0.013)	(0.047)
Dependency Ratio	-0.197***	-0.174***	-0.148***			-0.599***	-0.096***	-0.213***	-
	(0.041)	(0.040)	(0.042)			(0.135)	(0.026)	(0.045)	(0.153)
# of HH Members				-0.007***					
				(0.002)					
# of Children					-0.018***				
					(0.004)				
Mountain Treatment	-0.002	-0.003	-0.003	-0.009	-0.006	-0.171	-0.002	-0.032	0.005
	(0.027)	(0.028)	(0.027)	(0.028)	(0.027)	(0.124)	(0.013)	(0.045)	(0.123)
Coastal Treatment	0.135***	0.153***	0.149***	0.138***	0.135***	0.117	0.064***	0.073	0.165***
	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)	(0.203)	(0.021)	(0.065)	(0.203)
Coastal Control	0.220***	0.228***	0.226***	0.219***	0.217***	0.744***	0.099***		
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.106)	(0.025)		
Water Tank present	0.004	0.012	0.012	0.014	0.012	0.024	0.007	0.006	0.022
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.078)	(0.012)	(0.025)	(0.097)
TV present	0.040	0.043*	0.039	0.049**	0.046*				
	(0.025)	(0.025)	(0.025)	(0.025)	(0.024)				
TV and/or PC present									0.025
									(0.102)
Web Access	0.028	0.036	0.040	0.026	0.028	0.043	0.014	-0.001	
	(0.082)	(0.081)	(0.082)	(0.082)	(0.082)	(0.268)	(0.041)	(0.069)	
Edu. Durat° HH Head	0.009***							0.009***	
	(0.002)							(0.002)	
Literacy HH Head		0.043**							
		(0.021)							
Mean Edu. Durat° Adults			0.083						
			(0.051)						
Primary Edu				-0.008	-0.009	-0.024	-0.004		0.021
				(0.024)	(0.024)	(0.075)	(0.011)		(0.084)
Intermediate Edu				0.039	0.041	0.168	0.025		0.063
				(0.039)	(0.039)	(0.127)	(0.020)		(0.140)
Secondary Edu				0.064**	0.068**	0.225**	0.040**		0.108**
				(0.029)	(0.028)	(0.099)	(0.016)		(0.111)
Tertiary Edu				0.078**	0.083***	0.273**	0.050***		0.113**
				(0.032)	(0.031)	(0.111)	(0.018)		(0.125)
Sample Size	2466	2466	2470	2466	2466	2400	2466	2037	2037
Adj. R^2								0.114	
Pseudo R^2	0.094	0.087	0.086	0.086	0.090				0.102

Average marginal effects displayed for non-linear models. Robust standard errors in parentheses; bootstrapped for 2SPrB. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.17: Spending on soap, piped water, ICT covariates

Specification	(1)	(2)	(3)	(4)	(5)
Estimator	OLS	OLS	OLS	OLS	OLS
Piped Water	0.262***	0.224***	0.296***	0.264***	0.285***
	(0.073)	(0.073)	(0.073)	(0.074)	(0.073)
Age HH Head	-0.005**	-0.005**	-0.005*	-0.004*	-0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Gender HH Head	-0.113	-0.078	-0.072	-0.123	-0.070
	(0.180)	(0.182)	(0.182)	(0.184)	(0.182)
Marital Status HH Head	0.008	0.003	0.001	-0.019	-0.009
	(0.148)	(0.150)	(0.150)	(0.153)	(0.151)
Disease Awareness	0.037	0.030	0.061	0.038	0.059
	(0.055)	(0.056)	(0.056)	(0.056)	(0.056)
HH Member trained	-0.043	-0.023	-0.007	-0.033	0.013
	(0.138)	(0.141)	(0.138)	(0.136)	(0.136)
Asset Index	0.094**	0.102**	0.139***	0.049	0.127***
	(0.040)	(0.041)	(0.041)	(0.045)	(0.042)
Dependency Ratio	-0.515***	-0.534***	-0.501***	-0.479***	-0.490***
	(0.145)	(0.147)	(0.149)	(0.148)	(0.151)
Mountain Treatment	-0.075	-0.079	-0.109	-0.101	-0.096
	(0.077)	(0.078)	(0.076)	(0.077)	(0.076)
Coastal Treatment	-0.394***	-0.434***	-0.404***	-0.372***	-0.440***
	(0.106)	(0.107)	(0.107)	(0.107)	(0.106)
Coastal Control	-0.396***	-0.530***	-0.556***	-0.532***	-0.581***
	(0.101)	(0.104)	(0.107)	(0.104)	(0.106)
Water Tank present	0.191**	0.211**	0.224***	0.246***	0.223***
	(0.083)	(0.083)	(0.085)	(0.084)	(0.085)
TV present	0.819***				
	(0.103)				
Sat. Dish present		0.549***			
		(0.074)			
Radio present			0.202***		
			(0.059)		
GSM present				0.430***	
				(0.072)	
Landline Connection					0.231***
					(0.065)
Primary Edu	0.085	0.069	0.106	0.098	0.100
	(0.073)	(0.074)	(0.074)	(0.074)	(0.074)
Intermediate Edu	-0.199	-0.205	-0.166	-0.166	-0.167
	(0.129)	(0.129)	(0.131)	(0.130)	(0.132)
Secondary Edu	0.091	0.068	0.118	0.098	0.103
	(0.094)	(0.095)	(0.095)	(0.094)	(0.095)
Tertiary Edu	0.117	0.074	0.099	0.075	0.087
	(0.104)	(0.106)	(0.108)	(0.107)	(0.109)
Constant	3.287***	3.615***	3.707***	3.741***	3.827***
	(0.216)	(0.209)	(0.212)	(0.210)	(0.212)
Sample Size	2.262	2.262	2.262	2.262	2.262
Adj. R^2	0.113	0.100	0.076	0.089	0.076

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.18: Spending on soap, sewerage, ICT covariates

Specification	(1)	(2)	(3)	(4)	(5)
Estimator	OLS	OLS	OLS	OLS	OLS
Piped Water	0.277***	0.239***	0.297***	0.268***	0.290***
	(0.083)	(0.082)	(0.082)	(0.083)	(0.082)
Sewerage	-0.031	-0.030	-0.002	-0.009	-0.010
	(0.087)	(0.087)	(0.086)	(0.086)	(0.086)
Age HH Head	-0.005**	-0.005**	-0.005*	-0.004*	-0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Gender HH Head	-0.114	-0.079	-0.073	-0.123	-0.070
	(0.181)	(0.183)	(0.183)	(0.184)	(0.183)
Marital Status HH Head	0.009	0.005	0.001	-0.019	-0.009
	(0.148)	(0.150)	(0.150)	(0.154)	(0.151)
Disease Awareness	0.037	0.029	0.061	0.038	0.059
	(0.055)	(0.056)	(0.056)	(0.056)	(0.056)
HH Member trained	-0.040	-0.020	-0.007	-0.032	0.014
	(0.139)	(0.142)	(0.139)	(0.137)	(0.137)
Asset Index	0.094**	0.103**	0.139***	0.049	0.127***
	(0.040)	(0.041)	(0.041)	(0.045)	(0.042)
Dependency Ratio	-0.519***	-0.537***	-0.501***	-0.480***	-0.491***
	(0.146)	(0.148)	(0.150)	(0.149)	(0.151)
Mountain Treatment	-0.070	-0.074	-0.109	-0.100	-0.095
	(0.078)	(0.079)	(0.077)	(0.078)	(0.077)
Coastal Treatment	-0.381***	-0.420***	-0.404***	-0.368***	-0.436***
	(0.114)	(0.115)	(0.115)	(0.115)	(0.114)
Coastal Control	-0.392***	-0.527***	-0.556***	-0.531***	-0.580***
	(0.102)	(0.104)	(0.107)	(0.105)	(0.107)
Water Tank present	0.192**	0.213**	0.224***	0.246***	0.224***
	(0.083)	(0.084)	(0.085)	(0.084)	(0.085)
TV present	0.820***				
	(0.103)				
Sat. Dish present		0.550***			
		(0.074)			
Radio present			0.202***		
			(0.059)		
GSM present				0.430***	
				(0.073)	
Landline Connection					0.231***
					(0.065)
Primary Edu	0.085	0.069	0.106	0.098	0.100
	(0.073)	(0.074)	(0.074)	(0.074)	(0.074)
Intermediate Edu	-0.198	-0.204	-0.166	-0.166	-0.166
	(0.129)	(0.130)	(0.132)	(0.131)	(0.132)
Secondary Edu	0.092	0.069	0.118	0.098	0.103
	(0.094)	(0.095)	(0.095)	(0.094)	(0.095)
Tertiary Edu	0.118	0.074	0.099	0.075	0.087
	(0.104)	(0.106)	(0.108)	(0.107)	(0.109)
Constant	3.282***	3.611***	3.707***	3.740***	3.825***
	(0.216)	(0.209)	(0.211)	(0.209)	(0.212)
Sample Size	2262	2262	2263	2262	2262
Adj. R^2	0.113	0.100	0.072	0.088	0.076

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.19: Spending on soap, piped water, other specifications & estimators

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	2SLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	0.240	0.251***	0.262***	0.270***	0.273***	0.291***	0.303***	0.271***
	(0.173)	(0.059)	(0.073)	(0.073)	(0.073)	(0.073)	(0.071)	(0.072)
Age HH Head	-0.004	-0.005**	-0.005**	-0.005*	-0.005**	-0.005**	-0.001	-0.006**
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Gender HH Head	-0.134	-0.114	-0.113	-0.112	-0.114	-0.095	0.023	-0.073
	(0.179)	(0.180)	(0.180)	(0.179)	(0.177)	(0.177)	(0.177)	(0.179)
Marital Status HH Head	0.014	0.006	0.008	0.017	0.007	0.020	-0.007	0.052
	(0.147)	(0.148)	(0.148)	(0.147)	(0.147)	(0.146)	(0.143)	(0.145)
Disease Awareness		0.039	0.037	0.035	0.035	0.042	0.036	0.037
		(0.055)	(0.055)	(0.055)	(0.055)	(0.056)	(0.055)	(0.055)
Symptom Occurrence	0.089							
	(0.058)							
HH Member trained	-0.043	-0.042	-0.040	-0.029	-0.027	-0.013	-0.016	-0.027
	(0.149)	(0.140)	(0.140)	(0.140)	(0.140)	(0.142)	(0.139)	(0.139)
Asset Index	0.094**	0.095**	0.095**	0.097**	0.096**	0.114***	0.143***	0.119***
	(0.041)	(0.041)	(0.041)	(0.041)	(0.040)	(0.041)	(0.041)	(0.040)
Dependency Ratio	-0.547***	-0.519***	-0.516***	-0.515***	-0.509***	-0.558***		
	(0.154)	(0.145)	(0.145)	(0.145)	(0.145)	(0.151)		
# of HH Members							-0.060***	
							(0.008)	
# of Children								-0.089***
								(0.014)
Region	0.326***	0.338***						
	(0.066)	(0.063)						
Mountain Treatment			-0.075	-0.079	-0.085	-0.076	-0.105	-0.087
			(0.077)	(0.077)	(0.077)	(0.077)	(0.076)	(0.076)
Coastal Treatment			-0.394***	-0.398***	-0.396***	-0.366***	-0.461***	-0.443***
			(0.106)	(0.106)	(0.106)	(0.107)	(0.105)	(0.106)
Coastal Control			-0.396***	-0.396***	-0.392***	-0.392***	-0.439***	-0.424***
			(0.101)	(0.101)	(0.101)	(0.101)	(0.101)	(0.101)
Water Tank present	0.196**	0.189**	0.190**	0.190**	0.190**	0.207**	0.220***	0.201**
	(0.084)	(0.083)	(0.083)	(0.083)	(0.083)	(0.082)	(0.082)	(0.083)
TV present	0.812***	0.822***	0.818***	0.816***	0.815***	0.823***	0.860***	0.831***
	(0.111)	(0.103)	(0.103)	(0.103)	(0.104)	(0.103)	(0.103)	(0.103)
Web Access	-0.056	-0.064	-0.067	-0.068	-0.069	-0.055	-0.120	-0.095
	(0.313)	(0.316)	(0.315)	(0.313)	(0.312)	(0.308)	(0.303)	(0.312)
Edu. Durat° HH Head	0.006			0.006				
	(0.007)			(0.006)				
Literacy HH Head					0.090			
					(0.070)			
Mean Edu. Durat° Adults						-0.216		
						(0.175)		
Primary Edu		0.084	0.085				0.084	0.081
		(0.072)	(0.073)				(0.072)	(0.073)
Intermediate Edu		-0.201	-0.198				-0.247*	-0.222*
		(0.129)	(0.129)				(0.128)	(0.129)
Secondary Edu		0.091	0.091				0.027	0.069
		(0.093)	(0.094)				(0.093)	(0.093)
Tertiary Edu		0.118	0.118				0.044	0.089
		(0.103)	(0.104)				(0.103)	(0.103)
Constant	2.889***	2.896***	3.286***	3.251***	3.242***	4.221***	3.128***	3.230***
	(0.207)	(0.204)	(0.217)	(0.216)	(0.216)	(0.796)	(0.205)	(0.207)
Sample Size	2211	2262	2262	2262	2260	2263	2262	2262
Adj. R^2	0.110	0.113	0.113	0.112	0.112	0.112	0.128	0.123

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.20: Spending on soap, sewerage, other specifications & estimators

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	2SLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	0.574	0.273***	0.277***	0.287***	0.289***	0.299***	0.321***	0.296***
	(0.468)	(0.078)	(0.083)	(0.083)	(0.083)	(0.083)	(0.082)	(0.082)
Sewerage	-0.247	-0.033	-0.031	-0.035	-0.034	-0.018	-0.036	-0.051
	(0.335)	(0.083)	(0.087)	(0.087)	(0.087)	(0.087)	(0.086)	(0.086)
Age HH Head	-0.005*	-0.005**	-0.005**	-0.005*	-0.005**	-0.005**	-0.001	-0.006**
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Gender HH Head	-0.128	-0.116	-0.114	-0.113	-0.115	-0.095	0.022	-0.075
	(0.179)	(0.181)	(0.181)	(0.179)	(0.178)	(0.177)	(0.177)	(0.180)
Marital Status HH Head	0.026	0.008	0.009	0.018	0.009	0.020	-0.006	0.054
	(0.148)	(0.149)	(0.148)	(0.148)	(0.147)	(0.146)	(0.143)	(0.146)
Disease Awareness		0.039	0.037	0.034	0.034	0.042	0.036	0.037
		(0.055)	(0.055)	(0.055)	(0.055)	(0.056)	(0.055)	(0.055)
Symptom Occurrence	0.071							
	(0.062)							
HH Member trained	-0.067	-0.039	-0.038	-0.027	-0.025	-0.011	-0.013	-0.022
	(0.147)	(0.141)	(0.141)	(0.141)	(0.140)	(0.143)	(0.140)	(0.139)
Asset Index	0.102**	0.095**	0.095**	0.098**	0.097**	0.114***	0.144***	0.120***
	(0.043)	(0.041)	(0.041)	(0.041)	(0.040)	(0.041)	(0.040)	(0.040)
Dependency Ratio	-0.531***	-0.523***	-0.519***	-0.518***	-0.512***	-0.559***		
	(0.151)	(0.145)	(0.146)	(0.146)	(0.145)	(0.152)		
# of HH Members							(0.082)	
							-0.036	
# of Children								-0.089***
								(0.014)
Region	0.310***	0.333***						
	(0.069)	(0.065)						
Mountain Treatment			-0.071	-0.074	-0.080	-0.073	-0.100	-0.080
			(0.078)	(0.079)	(0.079)	(0.079)	(0.077)	(0.078)
Coastal Treatment			-0.381***	-0.383***	-0.381***	-0.358***	-0.445***	-0.421***
			(0.114)	(0.114)	(0.114)	(0.114)	(0.113)	(0.113)
Coastal Control			-0.393***	-0.393***	-0.389***	-0.390***	-0.435***	-0.419***
			(0.102)	(0.102)	(0.102)	(0.102)	(0.101)	(0.101)
Water Tank present	0.213**	0.190**	0.192**	0.191**	0.192**	0.208**	0.222***	0.203**
	(0.087)	(0.083)	(0.083)	(0.083)	(0.083)	(0.082)	(0.082)	(0.083)
TV present	0.785***	0.823***	0.819***	0.817***	0.816***	0.824***	0.861***	0.832***
	(0.111)	(0.103)	(0.103)	(0.103)	(0.104)	(0.103)	(0.103)	(0.103)
Web Access	-0.077	-0.063	-0.066	-0.067	-0.068	-0.054	-0.118	-0.093
	(0.314)	(0.316)	(0.315)	(0.312)	(0.312)	(0.308)	(0.303)	(0.311)
Edu. Durat° HH Head	0.004			0.006				
	(0.007)			(0.006)				
Literacy HH Head					0.091			
					(0.070)			
Mean Edu. Durat° Adults						-0.213		
						(0.176)		
Primary Edu		0.084	0.085				0.084	0.081
		(0.072)	(0.073)				(0.072)	(0.073)
Intermediate Edu		-0.199	-0.197				-0.246*	-0.220*
		(0.129)	(0.129)				(0.128)	(0.129)
Secondary Edu		0.093	0.092				0.028	0.070
		(0.093)	(0.094)				(0.093)	(0.093)
Tertiary Edu		0.120	0.119				0.045	0.090
		(0.103)	(0.104)				(0.103)	(0.103)
Constant	2.848***	0.084	3.281***	3.246***	3.237***	4.206***	0.084	3.221***
	(0.209)	(0.072)	(0.216)	(0.215)	(0.215)	(0.797)	(0.072)	(0.206)
Sample Size	2211	2262	2262	2262	2260	2263	2262	2262
Adj. R^2	0.104	0.113	0.112	0.111	0.111	0.112	0.128	0.123

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.21: Spending on cleanser, piped water, ICT covariates

Specification	(1)	(2)	(3)	(4)	(5)
Estimator	OLS	OLS	OLS	OLS	OLS
Piped Water	0.070	0.032	0.130*	0.049	0.113
	(0.074)	(0.075)	(0.073)	(0.075)	(0.074)
Age HH Head	-0.010***	-0.010***	-0.009***	-0.009***	-0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Gender HH Head	-0.090	-0.053	-0.044	-0.107	-0.044
	(0.180)	(0.186)	(0.190)	(0.187)	(0.190)
Marital Status HH Head	0.106	0.093	0.086	0.051	0.080
	(0.146)	(0.151)	(0.155)	(0.154)	(0.155)
Disease Awareness	0.213***	0.215***	0.254***	0.217***	0.249***
	(0.059)	(0.060)	(0.061)	(0.060)	(0.061)
HH Member trained	0.344***	0.376***	0.405***	0.357***	0.414***
	(0.104)	(0.106)	(0.102)	(0.100)	(0.100)
Asset Index	0.100**	0.122***	0.179***	0.034	0.158***
	(0.043)	(0.044)	(0.045)	(0.049)	(0.046)
Dependency Ratio	-0.586***	-0.594***	-0.581***	-0.537***	-0.556***
	(0.159)	(0.162)	(0.166)	(0.164)	(0.167)
Mountain Treatment	0.015	0.008	-0.020	-0.008	-0.016
	(0.076)	(0.077)	(0.073)	(0.076)	(0.073)
Coastal Treatment	-0.316***	-0.360***	-0.361***	-0.256**	-0.376***
	(0.107)	(0.109)	(0.109)	(0.108)	(0.109)
Coastal Control	-0.388***	-0.559***	-0.628***	-0.551***	-0.631***
	(0.106)	(0.111)	(0.116)	(0.111)	(0.115)
Water Tank present	0.027	0.060	0.082	0.107	0.078
	(0.087)	(0.089)	(0.091)	(0.089)	(0.091)
TV present	1.068***				
	(0.120)				
Sat. Dish present		0.647***			
		(0.084)			
Radio present			0.099		
			(0.063)		
GSM present				0.608***	
				(0.080)	
Landline Connection					0.229***
					(0.070)
Primary Edu	0.016	0.014	0.042	0.035	0.032
	(0.078)	(0.080)	(0.081)	(0.081)	(0.081)
Intermediate Edu	-0.258*	-0.252*	-0.217	-0.222	-0.223
	(0.145)	(0.145)	(0.148)	(0.145)	(0.148)
Secondary Edu	-0.017	-0.026	0.021	-0.005	-0.000
	(0.100)	(0.103)	(0.103)	(0.101)	(0.103)
Tertiary Edu	-0.016	-0.058	-0.041	-0.086	-0.065
	(0.112)	(0.116)	(0.120)	(0.117)	(0.121)
Constant	4.152***	4.575***	4.753***	4.749***	4.844***
	(0.243)	(0.234)	(0.239)	(0.235)	(0.240)
Sample Size	1992	1992	1992	1992	1992
Adj. R^2	0.146	0.115	0.079	0.110	0.082

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.22: Spending on cleanser, sewerage, ICT covariates

Specification	(1)	(2)	(3)	(4)	(5)
Estimator	OLS	OLS	OLS	OLS	OLS
Piped Water	0.080	0.035	0.114	0.034	0.102
	(0.084)	(0.084)	(0.083)	(0.084)	(0.083)
Sewerage	-0.018	-0.005	0.030	0.030	0.022
	(0.085)	(0.087)	(0.086)	(0.085)	(0.086)
Age HH Head	-0.010***	-0.010***	-0.009***	-0.009***	-0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Gender HH Head	-0.091	-0.053	-0.042	-0.106	-0.043
	(0.181)	(0.186)	(0.191)	(0.188)	(0.190)
Marital Status HH Head	0.107	0.094	0.084	0.050	0.079
	(0.147)	(0.152)	(0.155)	(0.155)	(0.156)
Disease Awareness	0.213***	0.215***	0.254***	0.217***	0.249***
	(0.059)	(0.060)	(0.061)	(0.060)	(0.061)
HH Member trained	0.346***	0.376***	0.403***	0.355***	0.412***
	(0.104)	(0.107)	(0.102)	(0.101)	(0.100)
Asset Index	0.100**	0.122***	0.179***	0.034	0.158***
	(0.043)	(0.044)	(0.045)	(0.049)	(0.046)
Dependency Ratio	-0.588***	-0.595***	-0.578***	-0.535***	-0.553***
	(0.159)	(0.163)	(0.166)	(0.164)	(0.168)
Mountain Treatment	0.017	0.008	-0.023	-0.011	-0.018
	(0.076)	(0.078)	(0.074)	(0.077)	(0.074)
Coastal Treatment	-0.309***	-0.358***	-0.372***	-0.267**	-0.385***
	(0.113)	(0.116)	(0.116)	(0.114)	(0.115)
Coastal Control	-0.386***	-0.558***	-0.630***	-0.554***	-0.632***
	(0.106)	(0.111)	(0.116)	(0.112)	(0.115)
Water Tank present	0.028	0.060	0.081	0.105	0.077
	(0.088)	(0.089)	(0.091)	(0.090)	(0.092)
TV present	1.069***				
	(0.120)				
Sat. Dish present		0.647***			
		(0.084)			
Radio present			0.099		
			(0.063)		
GSM present				0.608***	
				(0.080)	
Landline Connection					0.229***
					(0.070)
Primary Edu	0.016	0.014	0.042	0.035	0.032
	(0.078)	(0.080)	(0.081)	(0.081)	(0.081)
Intermediate Edu	-0.257*	-0.251*	-0.219	-0.224	-0.225
	(0.146)	(0.146)	(0.148)	(0.146)	(0.149)
Secondary Edu	-0.016	-0.026	0.020	-0.006	-0.001
	(0.100)	(0.103)	(0.103)	(0.101)	(0.103)
Tertiary Edu	-0.016	-0.058	-0.042	-0.087	-0.065
	(0.112)	(0.116)	(0.120)	(0.117)	(0.121)
Constant	4.149***	4.574***	4.757***	4.753***	4.847***
	(0.243)	(0.234)	(0.239)	(0.235)	(0.240)
Sample Size	1992	1992	1992	1992	1992
Adj. R^2	0.146	0.115	0.078	0.110	0.081

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.23: Spending on cleanser, piped water, other specifications & estimators

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	2SLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	0.045	0.113*	0.073	0.080	0.081	0.088	0.124*	0.094
	(0.177)	(0.064)	(0.074)	(0.075)	(0.075)	(0.076)	(0.073)	(0.073)
Age HH Head	-0.008***	-0.010***	-0.010***	-0.009***	-0.009***	-0.009***	-0.006**	-0.011***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Gender HH Head	-0.120	-0.094	-0.089	-0.092	-0.079	-0.087	0.043	-0.044
	(0.180)	(0.179)	(0.180)	(0.179)	(0.177)	(0.176)	(0.180)	(0.180)
Marital Status HH Head	0.109	0.108	0.106	0.108	0.102	0.105	0.064	0.136
	(0.146)	(0.146)	(0.146)	(0.145)	(0.145)	(0.144)	(0.142)	(0.144)
Disease Awareness		0.214***	0.214***	0.213***	0.212***	0.214***	0.210***	0.213***
		(0.059)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)
Symptom Occurrence	0.174***							
	(0.060)							
HH Member trained	0.363***	0.356***	0.354***	0.363***	0.363***	0.374***	0.383***	0.372***
	(0.117)	(0.106)	(0.106)	(0.106)	(0.106)	(0.107)	(0.102)	(0.103)
Asset Index	0.098**	0.102**	0.103**	0.102**	0.104**	0.110**	0.146***	0.128***
	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.045)	(0.044)	(0.044)
Dependency Ratio	-0.639***	-0.592***	-0.587***	-0.587***	-0.583***	-0.638***		
	(0.167)	(0.157)	(0.158)	(0.158)	(0.158)	(0.164)		
# of HH Members							-0.055***	
							(0.009)	
# of Children								-0.091***
								(0.014)
Region	0.340***	0.362***						
	(0.069)	(0.065)						
Mountain Treatment			0.013	0.010	0.007	0.008	-0.031	-0.005
			(0.076)	(0.076)	(0.076)	(0.076)	(0.075)	(0.074)
Coastal Treatment			-0.317***	-0.317***	-0.318***	-0.308***	-0.399***	-0.376***
			(0.107)	(0.107)	(0.107)	(0.107)	(0.108)	(0.106)
Coastal Control			-0.389***	-0.383***	-0.382***	-0.385***	-0.444***	-0.420***
			(0.106)	(0.106)	(0.106)	(0.106)	(0.107)	(0.106)
Water Tank present	0.036	0.023	0.027	0.026	0.027	0.033	0.061	0.040
	(0.088)	(0.087)	(0.087)	(0.087)	(0.087)	(0.087)	(0.087)	(0.087)
TV present	1.077***	1.071***	1.065***	1.063***	1.069***	1.068***	1.103***	1.074***
	(0.129)	(0.120)	(0.120)	(0.120)	(0.121)	(0.120)	(0.121)	(0.121)
Web Access	-0.162	-0.202	-0.201	-0.206	-0.204	-0.204	-0.258	-0.240
	(0.350)	(0.353)	(0.353)	(0.349)	(0.349)	(0.345)	(0.334)	(0.347)
Edu. Durat° HH Head	0.001			-0.001				
	(0.007)			(0.006)				
Literacy HH Head					-0.026			
					(0.075)			
Mean Edu. Durat° Adults						-0.169		
						(0.164)		
Primary Edu		0.018	0.016				0.008	0.009
		(0.078)	(0.078)				(0.078)	(0.078)
Intermediate Edu		-0.252*	-0.256*				-0.311**	-0.284*
		(0.146)	(0.146)				(0.145)	(0.145)
Secondary Edu		-0.012	-0.016				-0.083	-0.044
		(0.100)	(0.100)				(0.100)	(0.100)
Tertiary Edu		-0.008	-0.013				-0.095	-0.049
		(0.112)	(0.112)				(0.112)	(0.112)
Constant	3.828***	3.783***	4.147***	4.107***	4.106***	4.834***	3.985***	4.081***
	(0.233)	(0.233)	(0.243)	(0.242)	(0.241)	(0.731)	(0.234)	(0.235)
Sample Size	1952	1992	1992	1992	1991	1993	1992	1992
Adj. R^2	0.141	0.147	0.146	0.146	0.146	0.146	0.155	0.155

Robust standard errors in parentheses. Levels of statistical significance: *

Table A.24: Spending on cleanser, sewerage, other specifications & estimators

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	2SLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	0.484	0.117	0.082	0.093	0.092	0.094	0.133	0.110
	(0.516)	(0.080)	(0.084)	(0.085)	(0.084)	(0.085)	(0.083)	(0.083)
Sewerage	-0.307	-0.006	-0.018	-0.024	-0.022	-0.012	-0.017	-0.031
	(0.382)	(0.083)	(0.085)	(0.085)	(0.085)	(0.086)	(0.085)	(0.085)
Age HH Head	-0.009***	-0.010***	-0.010***	-0.009***	-0.009***	-0.009***	-0.006**	-0.011***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Gender HH Head	-0.120	-0.095	-0.090	-0.093	-0.080	-0.087	0.042	-0.046
	(0.181)	(0.180)	(0.181)	(0.180)	(0.178)	(0.177)	(0.181)	(0.180)
Marital Status HH Head	0.126	0.108	0.107	0.109	0.103	0.105	0.065	0.137
	(0.149)	(0.147)	(0.147)	(0.146)	(0.146)	(0.145)	(0.143)	(0.145)
Disease Awareness		0.214***	0.214***	0.213***	0.212***	0.214***	0.210***	0.212***
		(0.059)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)
Symptom Occurrence	0.149**							
	(0.067)							
HH Member trained	0.328***	0.357***	0.355***	0.365***	0.365***	0.375***	0.384***	0.374***
	(0.115)	(0.107)	(0.107)	(0.106)	(0.106)	(0.107)	(0.103)	(0.104)
Asset Index	0.110**	0.102**	0.104**	0.103**	0.104**	0.110**	0.147***	0.129***
	(0.046)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)
Dependency Ratio	-0.613***	-0.593***	-0.589***	-0.589***	-0.585***	-0.639***		
	(0.164)	(0.158)	(0.159)	(0.159)	(0.158)	(0.164)		
# of HH Members							-0.055***	
							(0.009)	
# of Children								-0.091***
								(0.015)
Region	0.335***	0.361***						
	(0.069)	(0.067)						
Mountain Treatment			0.015	0.012	0.010	0.009	-0.030	-0.002
			(0.077)	(0.077)	(0.077)	(0.077)	(0.076)	(0.075)
Coastal Treatment			-0.310***	-0.308***	-0.310***	-0.304***	-0.393***	-0.365***
			(0.113)	(0.113)	(0.113)	(0.113)	(0.113)	(0.112)
Coastal Control			-0.387***	-0.381***	-0.381***	-0.384***	-0.443***	-0.418***
			(0.106)	(0.106)	(0.106)	(0.106)	(0.107)	(0.106)
Water Tank present	0.060	0.023	0.028	0.027	0.028	0.033	0.062	0.042
	(0.093)	(0.088)	(0.088)	(0.088)	(0.087)	(0.087)	(0.087)	(0.088)
TV present	1.039***	1.072***	1.066***	1.063***	1.070***	1.068***	1.104***	1.075***
	(0.128)	(0.120)	(0.120)	(0.120)	(0.121)	(0.120)	(0.121)	(0.121)
Web Access	-0.198	-0.202	-0.201	-0.206	-0.204	-0.204	-0.258	-0.240
	(0.350)	(0.353)	(0.353)	(0.349)	(0.349)	(0.345)	(0.334)	(0.347)
Edu. Durat° HH Head	-0.003			-0.001				
	(0.007)			(0.006)				
Literacy HH Head					-0.025			
					(0.075)			
Mean Edu. Durat° Adults						-0.168		
						(0.165)		
Primary Edu		0.018	0.017				0.008	0.009
		(0.078)	(0.078)				(0.078)	(0.078)
Intermediate Edu		-0.252*	-0.255*				-0.310**	-0.282*
		(0.146)	(0.146)				(0.146)	(0.146)
Secondary Edu		-0.012	-0.016				-0.082	-0.043
		(0.100)	(0.100)				(0.100)	(0.100)
Tertiary Edu		-0.008	-0.012				-0.095	-0.048
		(0.112)	(0.112)				(0.112)	(0.112)
Constant	3.771***	3.784***	4.145***	4.104***	4.103***	4.824***	3.982***	4.076***
	(0.236)	(0.233)	(0.244)	(0.242)	(0.242)	(0.737)	(0.235)	(0.236)
Sample Size	1952	1992	1992	1992	1991	1993	1992	1992
Adj. R^2	0.133	0.146	0.146	0.145	0.145	0.146	0.155	0.154

Robust standard errors in parentheses. Levels of statistical significance: *

Table A.25: Latrine-cleaning frequency, piped water, ICT covariates

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	0.435*	0.421*	0.441*	0.428*	0.420*	0.439*
	(0.226)	(0.227)	(0.225)	(0.226)	(0.225)	(0.224)
Age HH Head	0.002	0.002	0.002	0.002	0.001	0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Gender HH Head	0.115	0.116	0.112	0.106	0.119	0.115
	(0.308)	(0.308)	(0.309)	(0.309)	(0.309)	(0.308)
Marital Status HH Head	0.224	0.225	0.227	0.219	0.216	0.226
	(0.259)	(0.259)	(0.259)	(0.259)	(0.260)	(0.259)
Disease Awareness	-0.024	-0.031	-0.022	-0.030	-0.028	-0.023
	(0.139)	(0.139)	(0.138)	(0.139)	(0.139)	(0.138)
HH Member trained	0.914*	0.909*	0.926*	0.905*	0.918*	0.923*
	(0.489)	(0.488)	(0.487)	(0.489)	(0.490)	(0.477)
Asset Index	0.590***	0.581***	0.602***	0.566***	0.570***	0.595***
	(0.108)	(0.109)	(0.109)	(0.116)	(0.111)	(0.109)
Dependency Ratio	0.340	0.337	0.321	0.354	0.369	0.338
	(0.309)	(0.309)	(0.311)	(0.310)	(0.309)	(0.310)
Mountain Treatment	-0.339	-0.339	-0.331	-0.345	-0.340	-0.340
	(0.254)	(0.254)	(0.254)	(0.254)	(0.254)	(0.254)
Coastal Treatment	0.392	0.388	0.369	0.405	0.395	0.391
	(0.330)	(0.329)	(0.329)	(0.331)	(0.329)	(0.329)
Coastal Control	-0.377	-0.372	-0.408	-0.371	-0.370	-0.383
	(0.256)	(0.250)	(0.250)	(0.249)	(0.249)	(0.249)
Water Tank present	0.547***	0.545***	0.552***	0.551***	0.542***	0.548***
	(0.178)	(0.178)	(0.178)	(0.178)	(0.177)	(0.177)
Soap bought	0.287	0.261	0.303	0.273	0.278	0.293
	(0.199)	(0.198)	(0.196)	(0.197)	(0.196)	(0.195)
TV present	0.032					
	(0.201)					
Sat. Dish present		0.128				
		(0.153)				
Radio present			-0.107			
			(0.146)			
GSM present				0.108		
				(0.154)		
Landline Connection					0.172	
					(0.177)	
Web Access						-0.177
						(0.897)
Primary Edu	-0.056	-0.063	-0.053	-0.058	-0.060	-0.054
	(0.182)	(0.182)	(0.182)	(0.182)	(0.182)	(0.182)
Intermediate Edu	-0.091	-0.099	-0.085	-0.091	-0.095	-0.087
	(0.272)	(0.272)	(0.273)	(0.272)	(0.272)	(0.273)
Secondary Edu	-0.167	-0.179	-0.162	-0.172	-0.184	-0.165
	(0.213)	(0.214)	(0.214)	(0.214)	(0.214)	(0.214)
Tertiary Edu	0.330	0.320	0.340	0.320	0.306	0.334
	(0.272)	(0.273)	(0.272)	(0.273)	(0.272)	(0.268)
Constant	5.321***	5.329***	5.355***	5.344***	5.398***	5.329***
	(0.515)	(0.508)	(0.508)	(0.509)	(0.515)	(0.509)
Sample Size	2.319	2.319	2.319	2.319	2.319	2.319
Adj. R^2	0.045	0.045	0.045	0.045	0.045	0.045

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.26: Latrine-cleaning frequency, sewerage, ICT covariates

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	0.407	0.395	0.413	0.400	0.394	0.410
	(0.262)	(0.263)	(0.261)	(0.262)	(0.261)	(0.261)
Sewerage	0.059	0.054	0.059	0.059	0.055	0.061
	(0.217)	(0.216)	(0.217)	(0.217)	(0.217)	(0.216)
Age HH Head	0.002	0.002	0.002	0.002	0.001	0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Gender HH Head	0.117	0.119	0.115	0.108	0.122	0.117
	(0.308)	(0.308)	(0.308)	(0.309)	(0.308)	(0.308)
Marital Status HH Head	0.221	0.222	0.224	0.216	0.214	0.223
	(0.259)	(0.259)	(0.259)	(0.259)	(0.260)	(0.259)
Disease Awareness	-0.024	-0.030	-0.021	-0.029	-0.027	-0.022
	(0.139)	(0.139)	(0.139)	(0.140)	(0.139)	(0.138)
HH Member trained	0.909*	0.904*	0.920*	0.900*	0.913*	0.918*
	(0.487)	(0.487)	(0.486)	(0.488)	(0.488)	(0.475)
Asset Index	0.589***	0.580***	0.601***	0.565***	0.569***	0.594***
	(0.108)	(0.109)	(0.109)	(0.116)	(0.111)	(0.109)
Dependency Ratio	0.346	0.343	0.328	0.361	0.375	0.345
	(0.311)	(0.311)	(0.313)	(0.312)	(0.311)	(0.312)
Mountain Treatment	-0.347	-0.346	-0.340	-0.353	-0.347	-0.348
	(0.253)	(0.252)	(0.253)	(0.252)	(0.252)	(0.252)
Coastal Treatment	0.367	0.365	0.344	0.379	0.371	0.364
	(0.329)	(0.328)	(0.327)	(0.329)	(0.328)	(0.327)
Coastal Control	-0.382	-0.377	-0.414*	-0.376	-0.375	-0.389
	(0.256)	(0.249)	(0.249)	(0.249)	(0.249)	(0.248)
Water Tank present	0.545***	0.542***	0.549***	0.548***	0.539***	0.545***
	(0.179)	(0.178)	(0.178)	(0.178)	(0.178)	(0.178)
Soap bought	0.286	0.260	0.302	0.272	0.277	0.292
	(0.199)	(0.198)	(0.196)	(0.197)	(0.196)	(0.195)
TV present	0.030					
	(0.201)					
Sat. Dish present		0.126				
		(0.153)				
Radio present			-0.107			
			(0.146)			
GSM present				0.108		
				(0.154)		
Landline Connection					0.171	
					(0.177)	
Web Access						-0.180
						(0.896)
Primary Edu	-0.056	-0.064	-0.053	-0.059	-0.061	-0.055
	(0.182)	(0.182)	(0.182)	(0.182)	(0.182)	(0.182)
Intermediate Edu	-0.093	-0.101	-0.088	-0.094	-0.097	-0.090
	(0.273)	(0.273)	(0.273)	(0.273)	(0.273)	(0.274)
Secondary Edu	-0.169	-0.181	-0.165	-0.175	-0.186	-0.168
	(0.214)	(0.215)	(0.214)	(0.214)	(0.214)	(0.214)
Tertiary Edu	0.328	0.318	0.338	0.317	0.304	0.332
	(0.271)	(0.273)	(0.272)	(0.272)	(0.272)	(0.268)
Constant	5.330***	5.337***	5.364***	5.352***	5.405***	5.338***
	(0.514)	(0.508)	(0.508)	(0.509)	(0.514)	(0.508)
Sample Size	2.319	2.319	2.319	2.319	2.319	2.319
Adj. R^2	0.044	0.045	0.044	0.044	0.045	0.044

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Table A.27: Latrine-cleaning frequency, piped water, other specifications & estimators

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	2SLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	1.511***	0.721***	0.438*	0.439**	0.444*	0.380*	0.411*	0.434*
	(0.449)	(0.148)	(0.225)	(0.223)	(0.225)	(0.229)	(0.226)	(0.225)
Age HH Head	-0.001	0.002	0.002	0.002	0.001	0.000	-0.001	0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Gender HH Head	0.105	0.052	0.113	0.102	0.139	0.143	0.008	0.085
	(0.317)	(0.308)	(0.308)	(0.310)	(0.297)	(0.294)	(0.305)	(0.305)
Marital Status HH Head	0.255	0.233	0.226	0.238	0.232	0.206	0.222	0.200
	(0.265)	(0.258)	(0.259)	(0.259)	(0.255)	(0.254)	(0.257)	(0.260)
Disease Awareness	0.059	-0.019	-0.023	-0.024	-0.021	-0.046	-0.020	-0.021
	(0.147)	(0.139)	(0.138)	(0.138)	(0.138)	(0.137)	(0.138)	(0.138)
HH Member trained	0.732	0.935**	0.921*	0.913*	0.914*	0.834*	0.902*	0.913*
	(0.520)	(0.476)	(0.477)	(0.477)	(0.477)	(0.481)	(0.477)	(0.477)
Asset Index	0.623***	0.580***	0.593***	0.605***	0.611*	0.556*	0.556*	0.577*
	(0.111)	(0.110)	(0.110)	(0.110)	(0.110)	(0.108)	(0.110)	(0.110)
Dependency Ratio	0.501	0.269	0.338	0.364	0.371	0.667*		
	(0.328)	(0.309)	(0.310)	(0.312)	(0.308)	(0.316)		
# of HH Members							0.046*	
							(0.018)	
# of Children								0.055*
								(0.031)
Region	-0.142	-0.299*						
	(0.176)	(0.168)						
Mountain Treatment			-0.339	-0.341	-0.338	-0.352	-0.317	-0.331
			(0.254)	(0.253)	(0.253)	(0.253)	(0.256)	(0.255)
Coastal Treatment			0.392	0.406	0.414	0.325	0.448	0.421
			(0.330)	(0.332)	(0.328)	(0.325)	(0.333)	(0.331)
Coastal Control			-0.376	-0.371	-0.370	-0.372	-0.339	-0.357
			(0.256)	(0.256)	(0.256)	(0.256)	(0.259)	(0.257)
Water Tank present	0.528***	0.487***	0.546***	0.562***	0.567*	0.508*	0.527*	0.539*
	(0.179)	(0.176)	(0.178)	(0.177)	(0.176)	(0.177)	(0.177)	(0.177)
Soap bought	0.283	0.264	0.286	0.287	0.288	0.308	0.292	0.289
	(0.202)	(0.201)	(0.199)	(0.200)	(0.199)	(0.199)	(0.199)	(0.199)
TV present	-0.035	0.122	0.030	0.014	0.014	-0.021	0.000	0.022
	(0.229)	(0.201)	(0.202)	(0.202)	(0.202)	(0.202)	(0.201)	(0.202)
Web Access	-0.284	-0.174	-0.174	-0.115	-0.108	-0.127	-0.132	-0.164
	(0.908)	(0.892)	(0.898)	(0.907)	(0.910)	(0.915)	(0.903)	(0.899)
Edu. Durat° HH Head	-0.011			0.001				
	(0.016)			(0.014)				
Literacy HH Head					-0.060			
					(0.162)			
Mean Edu. Durat° Adults						1.071*		
						(0.372)		
Primary Edu		-0.040	-0.055				-0.051	-0.051
		(0.182)	(0.182)				(0.182)	(0.182)
Intermediate Edu		-0.059	-0.088				-0.050	-0.072
		(0.270)	(0.273)				(0.274)	(0.273)
Secondary Edu		-0.118	-0.166				-0.121	-0.151
		(0.212)	(0.214)				(0.211)	(0.211)
Tertiary Edu		0.391	0.335				0.386	0.353
		(0.267)	(0.268)				(0.267)	(0.268)
Constant	4.850***	5.316***	5.318***	5.249***	5.271*	0.563	5.413*	5.361*
	(0.493)	(0.467)	(0.515)	(0.510)	(0.512)	-1.695	(0.497)	(0.501)
Sample Size	2260	2319	2319	2319	2320	2322	2319	2319
Adj. R^2	0.034	0.040	0.044	0.044	0.044	0.047	0.046	0.045

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, **

Table A.28: Latrine-cleaning frequency, sewerage, other specifications & estimators

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	2SLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Piped Water	2.288*	0.613***	0.409	0.411	0.413	0.379	0.380	0.400
	(1.212)	(0.216)	(0.261)	(0.260)	(0.261)	(0.262)	(0.262)	(0.262)
Sewerage	-0.908	0.166	0.060	0.059	0.066	0.003	0.065	0.071
	(0.855)	(0.217)	(0.216)	(0.216)	(0.216)	(0.216)	(0.214)	(0.216)
Age HH Head	-0.002	0.002	0.002	0.002	0.001	0.000	-0.001	0.002
	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Gender HH Head	0.075	0.063	0.116	0.105	0.141	0.143	0.010	0.087
	(0.317)	(0.308)	(0.308)	(0.310)	(0.297)	(0.293)	(0.304)	(0.305)
Marital Status HH Head	0.309	0.223	0.223	0.235	0.230	0.206	0.220	0.197
	(0.276)	(0.258)	(0.259)	(0.259)	(0.255)	(0.254)	(0.257)	(0.260)
Disease Awareness	0.064	-0.017	-0.023	-0.023	-0.020	-0.046	-0.019	-0.020
	(0.151)	(0.139)	(0.138)	(0.138)	(0.138)	(0.138)	(0.138)	(0.138)
HH Member trained	0.768	0.919*	0.916*	0.908*	0.909*	0.834*	0.896*	0.906*
	(0.518)	(0.474)	(0.476)	(0.475)	(0.476)	(0.479)	(0.476)	(0.476)
Asset Index	0.644***	0.578***	0.592***	0.604***	0.610*	0.556*	0.554*	0.576*
	(0.114)	(0.110)	(0.110)	(0.110)	(0.110)	(0.108)	(0.110)	(0.110)
Dependency Ratio	0.436	0.290	0.345	0.371	0.378	0.668*		
	(0.323)	(0.312)	(0.312)	(0.315)	(0.310)	(0.317)		
# of HH Members							0.047*	
							(0.018)	
# of Children								0.056*
								(0.031)
Region	-0.270	-0.269						
	(0.178)	(0.168)						
Mountain Treatment			-0.348	-0.350	-0.348	-0.352	-0.326	-0.341
			(0.253)	(0.252)	(0.251)	(0.251)	(0.254)	(0.253)
Coastal Treatment			0.366	0.380	0.385	0.324	0.419	0.390
			(0.329)	(0.330)	(0.326)	(0.324)	(0.331)	(0.329)
Coastal Control			-0.382	-0.376	-0.377	-0.372	-0.345	-0.363
			(0.256)	(0.255)	(0.255)	(0.255)	(0.258)	(0.257)
Water Tank present	0.555***	0.482***	0.544***	0.560***	0.563*	0.508*	0.524*	0.536*
	(0.186)	(0.177)	(0.178)	(0.178)	(0.177)	(0.178)	(0.178)	(0.178)
Soap bought	0.280	0.263	0.285	0.286	0.287	0.308	0.291	0.288
	(0.205)	(0.201)	(0.199)	(0.200)	(0.199)	(0.200)	(0.199)	(0.199)
TV present	-0.026	0.112	0.028	0.012	0.012	-0.021	-0.002	0.020
	(0.233)	(0.200)	(0.202)	(0.202)	(0.202)	(0.202)	(0.201)	(0.201)
Web Access	-0.274	-0.182	-0.177	-0.118	-0.112	-0.127	-0.136	-0.168
	(0.911)	(0.892)	(0.897)	(0.905)	(0.908)	(0.914)	(0.902)	(0.898)
Edu. Durat° HH Head	-0.011			0.001				
	(0.017)			(0.014)				
Literacy HH Head					-0.062			
					(0.163)			
Mean Edu. Durat° Adults						1.070*		
						(0.373)		
Primary Edu		-0.043	-0.055				-0.052	-0.051
		(0.183)	(0.182)				(0.182)	(0.182)
Intermediate Edu		-0.069	-0.091				-0.053	-0.075
		(0.271)	(0.273)				(0.274)	(0.274)
Secondary Edu		-0.128	-0.169				-0.123	-0.153
		(0.213)	(0.214)				(0.211)	(0.212)
Tertiary Edu		0.381	0.332				0.385	0.351
		(0.266)	(0.268)				(0.267)	(0.267)
Constant	4.834***	5.304***	5.327***	5.258***	5.282*	0.565	5.427*	5.374*
	(0.506)	(0.467)	(0.515)	(0.509)	(0.511)	(1.701)	(0.497)	(0.500)
Sample Size	2260	2319	2319	2319	2320	2322	2319	2319
Adj. R^2	0.016	0.040	0.044	0.043	0.043	0.047	0.046	0.045

Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, **

Table A.29: IV estimator first stage results, connection to piped water and to improved sewerage

Piped water

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	OLS	OLS	Probit	Probit	OLS	OLS	Probit	Probit
IV Distance to Centre					-0.00026***	-0.00022***	-0.00025***	-0.00020***
					(0.000)	(0.000)	(0.000)	(0.000)
IV House on Rock	-0.16164***	-0.10562***	-0.16354***	-0.09920***	-0.10239***	-0.06984**	-0.09296***	-0.05227*
	(0.035)	(0.034)	(0.035)	(0.033)	(0.034)	(0.033)	(0.033)	(0.031)
IV Age of House	0.00141***	0.00078***	0.00175***	0.00109***	0.00090***	0.00049***	0.00111***	0.00070***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample Size	2419	2369	2419	2369	2354	2307	2354	2307
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
Adj. R^2	0.031	0.157			0.162	0.244		
Pseudo R^2			0.026	0.134			0.124	0.205
F	39.480	30.468			152.227	47.598		
Chi^2			86.386	438.098			402.396	652.672
Model p-value	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***

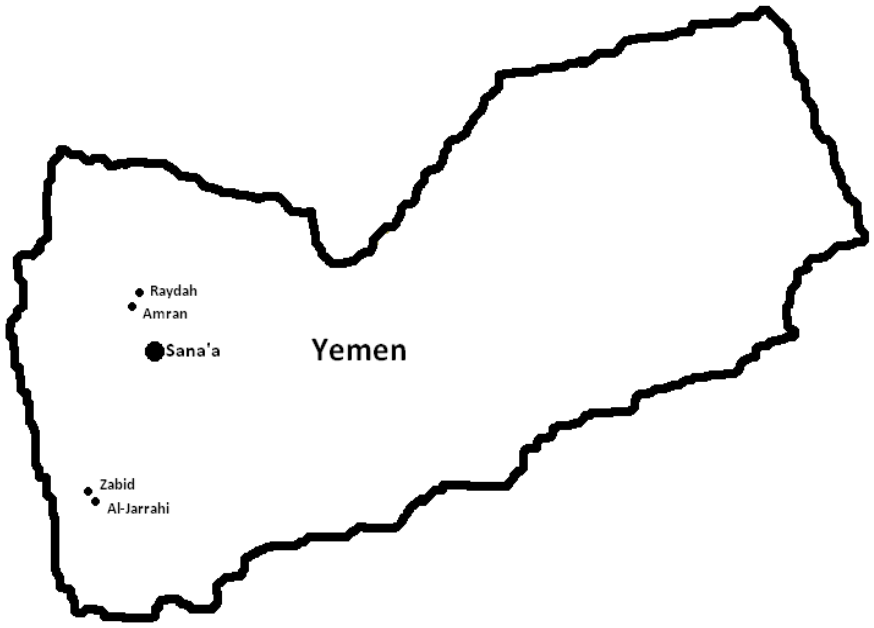
Improved sewerage

Specification	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Estimator	OLS	OLS	Probit	Probit	OLS	OLS	Probit	Probit
IV Distance to Centre					-0.00028***	-0.00024***	-0.00037***	-0.00027***
					(0.000)	(0.000)	(0.000)	(0.000)
IV House on Rock	-0.20633***	-0.15756***	-0.22479***	-0.16258***	-0.14258***	-0.11865***	-0.12721***	-0.11258***
	(0.035)	(0.034)	(0.037)	(0.035)	(0.033)	(0.032)	(0.036)	(0.035)
IV Age of House	0.00103***	0.00054***	0.00104***	0.00056***	0.00046***	0.00020	0.00027*	0.00014
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample Size	2419	2370	2419	2370	2354	2308	2354	2308
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
Adj. R^2	0.026	0.150			0.189	0.264		
Pseudo R^2			0.021	0.123			0.188	0.239
F	33.551	28.895			183.444	52.636		
Chi^2			68.441	395.012			596.654	745.484
Model p-value	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***

Probit coefficients are marginal effects. Note that the small sizes of the coefficients for IV Distance to Centre are rooted in the unit of measure of the variable in meters. An alternative measure in hundred metre increments (*i.e.*, shifting the decimal point by two places) yields coefficient sizes ranging between -0.037 and -0.020, always statistically significant at the one per cent level. Exemplary set of covariates used here comprises: Age HH Head, Gender HH Head, Marital Status HH Head, Disease Awareness, HH Member trained, Asset Index, Dependency Ratio, Water Tank present, TV present, and Primary to Tertiary Education Levels. Robust standard errors in parentheses. Levels of statistical significance: * p<0.10, ** p<0.05, *** p<0.01

Appendix 7: Geography

Figure A.1: Geographic location of survey towns



The map (author's drawing) shows the mountainous project (Amran) and control (Raydah) towns northward of the capital city Sana'a as well as those in the western coastal plain (Zabid and Al Jarrahi).